ERDC/CRREL TR-06-15

Cold Regions Research and Engineering Laboratory

US Army Corps of Engineers®
Engineer Research and Development Center

A Generalized Approach to Soil Strength Prediction with Machine Learning Methods

Peter M. Semen

July 2006

ERDC/CRREL TR-06-15
July 2006

A Generalized Approach to Soil Strength Prediction with Machine Learning Methods

Peter M. Semen

Cold Regions Research and Engineering Laboratory
U.S. Army Engineer Research and Development Center
72 Lyme Rd.
Hanover, NH 03755

Thayer School of Engineering
Dartmouth College

"A Generalized Approach to Soil Strength Prediction with Machine Learning Methods"

Peter M. Semen

Master of Science

Committee

Susan P. McGrath, Ph.D. (Chair)
Laura R. Ray, Ph.D.
Sally A. Shoop, Ph.D.

ABSTRACT

Current methods for evaluating the suitability of potential landing sites for fixed-wing aircraft require a direct measurement of soil bearing capacity. In contingency military operations, the commitment of ground troops to carry out this mission prior to landing poses problems in hostile territory, including logistics, safety, and operational security. Developments in remote sensing technology provide an opportunity to make indirect measurements that may prove useful for inferring basic soil properties. However, methods to accurately predict strength from other fundamental geotechnical parameters are lacking, especially for a broad range of soil types under widely-varying environmental conditions. To support the development of new procedures, a dataset of in situ soil pit test results was gathered from airfield pavement evaluations at forty-six locations worldwide that encompass a broad variety of soil types. Many features associated with soil strength—including gradation, moisture content, density, specific gravity and plasticity—were collected along with California bearing ratio (CBR), a critical strength index used to determine the traffic loading that the ground can support.

Machine learning methods—with advantages in nonlinear relationship mapping, nonparametric distribution treatment, superior generalization, and implicit modeling—were applied, hypothesizing these characteristics might make them better-suited to geotechnical problems. Artificial neural network and *k*-nearest neighbor techniques were tested on plastic and non-plastic subsets of data and compared to conventional regression and existing CBR prediction methods. The machine learning models were able to halve the baseline error rate for plastic soils, but non-plastic soils showed no significant improvement. For both groups, normalized root mean square error (NRMSE) for generalization to new cases was approximately fifty percent for the best models. The high degree of variability for direct soil strength measurement methods limits the lowest possible NRMSE to approximately twenty-five percent, even before introducing any additional errors expected with remote sensing.

Acknowledgement

I would like to thank the agencies that provided the funding that made this work possible. They include the US Transportation Command (TRANSCOM), the US Air Force, Air Mobility Command (AMC), the US Air Force, Air Force Research Laboratory (AFRL), and the US Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory (USACE-ERDC-CRREL). I am also grateful to CRREL for their generous support of my school tuition throughout this degree program.

I would like to thank the members of my examining committee for their help and guidance throughout the process. To my advisor, Professor Susan McGrath for her willingness to adopt an "orphaned" graduate student, coach me through the difficulties I encountered, and for her constant positive encouragement that kept me going. To Professor Laura Ray for her willingness to participate and provide valuable input, despite a heavy workload. And to Dr. Sally Shoop who championed the work throughout and had the confidence in me to entrust me with this work Her detailed feedback and patience with me during the process were invaluable and greatly appreciated.

I would also like to thank Dr. Shoop and Dr. Charles Ryerson for their careful oversight of the Opportune Landing Site program and making sure that this work remained a priority, despite budgetary pressures.

I would like to acknowledge the assistance from others that I received along the way in carrying out this work. Thanks to Ms. Deborah Diemand, Mr. James Cole-Henry, and Ms. Lynette Barna for their diligent efforts in accurately entering data for the database. Special thanks to Dr. Raymond Rollings for generously sharing his expertise and advice on all things geotechnical and providing valuable input on how to approach the problem along the way. His consolations regarding the difficulties that soil problems can pose were very comforting, coming from a true expert in the field. Also, thanks to Mr. Lawrence Gatto for his advice on the relative merits of using geomorphological features to predict soil types.

I would like to thank my parents for their lifelong support and encouragement, especially for my education. And finally, to my wife Jessica and our dog Cabot for their constant love and support at every step of the way. Without their patience and continuous supply of encouragement, none of this would have been possible.

Table of Contents

List of Tables

List of Illustrations

1 Introduction

Current methods for evaluating potential landing sites for fixed-wing aircraft require a direct measurement of bearing capacity be made prior to landing. The commitment of ground troops to carry out this mission poses many problems, including logistics, safety, and operational security. With limited resources to deploy, this approach is not practical for supporting the future strategic vision of using multiple unimproved access points, and even proves difficult in present-day situations where several candidate landing sites across a wide area need to be assessed. Safety and security of those involved in these evaluations are also a great concern, as the US Air Force has described in their latest planning documents. "The Mobility Air Force [MAF] personnel experience a high risk situation when deployed to potential austere landing sites to measure the bearing capacity of the soil. This bearing capacity must be determined before MAF aircraft can safely land. These site survey teams are exposed to danger and could compromise the landing site if discovered" [1].

A dramatic example of this situation occurred recently during Operation Enduring Freedom in Afghanistan. In a nighttime raid on a Taliban outpost, the structural capacity of an adjacent unsurfaced runway proved critical to extracting the forces after the raid and for subsequent use of the airstrip as a forward operating base (Figure 1.1). "While the [Army] Rangers secured the perimeter and searched the compound's buildings, specialists from the Air Force…walked the runway in the dark. Alert for mines, they tested the runway with a soil penetrometer, a long rod with a cone-shaped end and a sliding weight. It registers soil resistance when the weight is dropped, allowing bearing characteristics to be evaluated" [2]. The use of the dynamic cone penetrometer (DCP), a

handheld tool that can generate considerable noise when the sliding weight hits the anvil, could attract unwanted attention in such a situation. Despite these risks, the need to measure the soil strength was crucial. An officer who participated in the planning and execution of the raid stated, “The most important intelligence we brought back was really the condition of that runway….That was something we could not tell until we were actually on the ground and sampled it. We brought along specialists who walked up and down the runway to take readings and measurements” [2].

Figure 1.1: A Soldier Provides Perimeter Security for C-17 Operations on an Unsurfaced Airstrip at a Forward Operating Base in Afghanistan (Source: United States Air Force).

1.1 Opportune Landing Site Program

According to the latest military doctrine, strategic responsiveness and deployability of combat forces will be one of the key requirements for decisive

operations in future conflicts. An increasing desire to use unimproved access points, which allow commanders planning flexibility and preserve the element of surprise, is driving the demand for vehicles that can operate in austere conditions, with little or no infrastructure to support them. For airborne maneuvers, the US Training and Doctrine Command have outlined their operational needs for a Super Short Takeoff and Landing (SSTOL) aircraft in the following way: "This is joint [combined services] airlift with the ability to carry two light/medium armored vehicles 3500 miles. It can land on 750 feet of road or field in the joint area of operations, which avoids fixed airfields and adds innumerable points of entry. Its features provide the joint commander sharply improved options to employ mounted ground forces to achieve operational surprise and conduct air-ground maneuver throughout the Joint Operational Area (JOA)" [3].

Boeing's concept for this Advanced Theater Transport is a tailless turboprop engine aircraft dubbed the "Super Frog" (Figure 1.2) with the ability to tilt its wings for operating on unprepared runways as short as 650 to 1,000 feet [4]. Envisioned as a replacement for the current C-130, the tactical airlifter will have a larger fuselage cross-section that is comparable to the C-17 and an estimated cargo capacity of up to 80,000 pounds. Recognizing the need for a methodology to select appropriate landing zones for such planes, Boeing began an Opportune Landing Site (OLS) project in 1998. "One of the prime objectives of OLS has been to develop an automatic method that employs commercially accessible satellite data for finding the best available, un-improved landing sites for aircraft in natural terrain" [5].

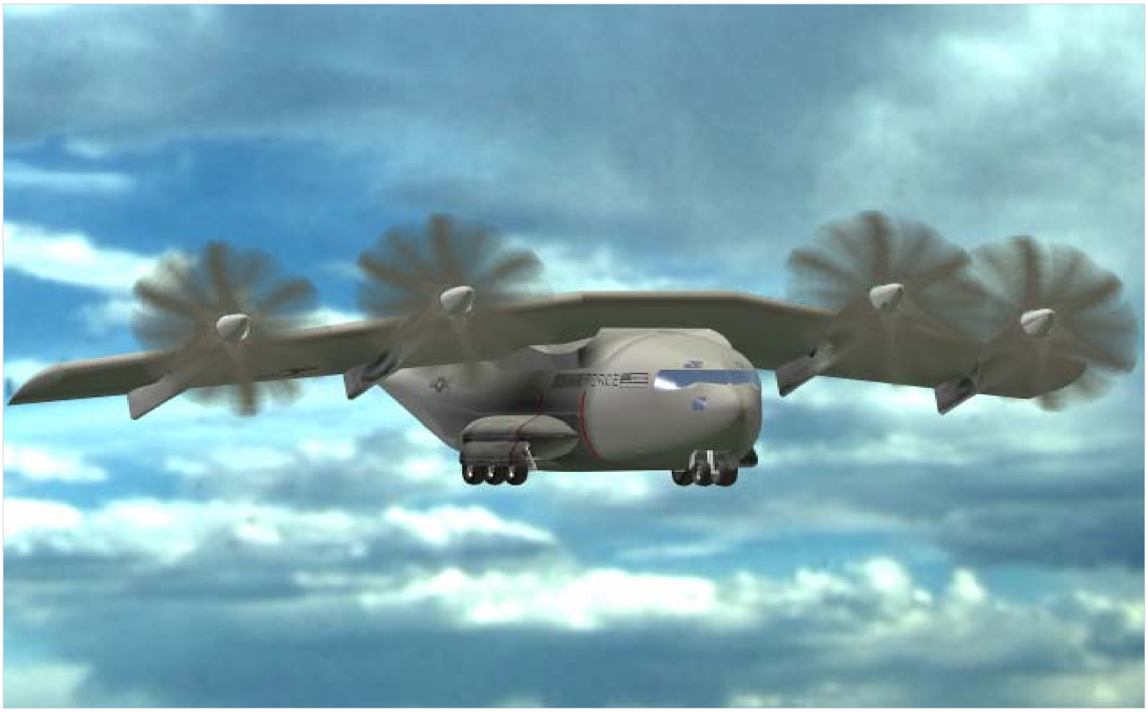

Figure 1.2: Conceptual Illustration of Boeing's Advanced Tactical Transport Aircraft for Strategic Military Airlift (Source: Boeing Phantom Works).

The method utilizes LANDSAT satellite data to locate areas that could be suitable for landing, offloading, loading, and takeoff of aircraft. Several criteria must be met for a site to be selected by the software. First, the landing site must be flat and long enough to meet the basic geometric requirements for an airstrip. The area must also be free from obstructions, such as powerlines, ditches, and roads. (In this instance, roads are not being considered as landing sites.) The candidate site must lack significant vegetation cover such as shrubs or trees that could interfere with operations and potentially damage the aircraft. The method also chooses areas that are free from standing surface water. The final element that must be satisfied is that the ground must be strong enough to support aircraft traffic loading. Reportedly, the development efforts for Boeing's OLS software have been focused on identifying areas that meet the topographic criteria and not directly

on soil strength, however the technique has shown some promise in locating firmer locations [5].

The US Army Corps of Engineers, Engineer Research and Development Center, Cold Regions Research and Engineering Laboratory (ERDC-CRREL) has been tasked with providing an independent evaluation of the software's performance for the US Air Force. Using areas selected by the software as suitable landing sites, detailed geotechnical characterization of each site and an evaluation regarding its suitability to land current military aircraft are underway. Sites are being tested during different seasons to determine seasonal variations in areas selected by the software.

In addition to providing an evaluation of the Boeing OLS software, the ERDC team is also working to develop better methods to predict soil strength. This study is a part of that effort whose goal is to document, develop, and refine relationships between soil strength and other soil parameters in order to allow the selection of OLS sites based on their structural suitability to support aircraft loading. Parallel work has been underway within the Joint Rapid Airfield Construction (JRAC) research program to provide military planners and warfighters with fast-track construction techniques to support quick-response deployments to areas where the existing infrastructure is lacking [6]. While the remote site assessment and material characterization tools being developed under the JRAC effort have been focused on providing engineering solutions to airfield construction, the soil strength prediction element of the OLS program is chiefly concerned with non-engineered, largely unimproved sites for immediate expedient or possible emergency use.

1.2 Pedotransfer Functions

The direct measurement of soil physical properties can be costly and time consuming. Therefore, the ability to estimate a soil property from other fundamental parameters that can be measured more easily and cheaply is useful and valuable. Although the use of such estimates and "rules of thumb" have existed for a long time, the formal term "pedotransfer function" has recently come into use by the soil science community to describe these relationships among geotechnical parameters that are able to "translate data we *have* to data we *need*" [7]. Recently, the major focus in pedotransfer function development has been in predicting the hydraulic properties of soil, though others have been developed for chemical, physical, mechanical, and biological properties [8,9]. Largely, this trend has been driven by the demands of numerical simulation models that require a large number of soil properties to run [8,9]. In the case of the OLS task, we seek relationships to predict the California bearing ratio* soil strength index (described in detail in section 2.1) from other physical properties and related index test values. There has been limited work in the soil science and geotechnical community in developing these relationships, so consequently "pedotransfer functions for soil strength are scarce" [10].

Traditionally, in terms of the numerical techniques used to develop these relationships, "the vast majority of pedotransfer functions are empirically based on linear regression equations with a few being physical model methods and neural network models" [8]. However, the past decade has seen strong growth in the use of machine learning and artificial intelligence techniques to develop new geotechnical models and

---

* Short definitions for terminology specific to geotechnical engineering can be found in a glossary in Appendix B.

pedotransfer functions [11,12]. Another novel approach that is actually based upon a rule-based system (a method detailed in section 4.2.2) is SINFERS, an initial attempt at developing a knowledge-based **s**oil **infer**ence **s**ystem where pedotransfer functions serve as the "rules" for an expert system [9]. With SINFERS, the goal is to link together many of the pedotransfer functions that have already been published into an integrated system where users can use the properties they know to estimate others that they require. The approach is rather unique because it incorporates the concept of membership, where predictions are limited to soils that are like the ones used to originally develop the pedotransfer function. It also begins to deal with the difficult problem of estimating the degree of uncertainty for different relationships, a concept that has unfortunately been neglected to a large extent in most other attempts to link soil properties.

1.3 Hypothesis

The goal of this work is to determine whether recent advancements in machine learning techniques can be used to improve our ability to predict the California bearing ratio (CBR) soil strength index. While there has been some success using these methods to predict CBR and other soil strength indices for specific soils in very limited geographic locations, application to a wide variety of soils representative of worldwide conditions has not been attempted. This study hypothesizes that emerging developments in adaptive learning methods may be more suited to estimating CBR values for a wide range of soil types and conditions than existing techniques based on traditional statistical methods.

Some characteristics of these techniques (discussed in section 4.2) appear well-suited to the objectives and constraints of the OLS soil strength prediction problem. These methods have advanced the analysis of other complex phenomena that do not lend

themselves well to explicit modeling—such as stock market analysis, medical diagnosis, drug discovery, credit risk estimation, and weather forecasting [13]. Consequently, there is reason to believe that these techniques may also be promising in modeling other difficult systems, such as characterizing soil behavior.

The final intention of this work is to provide a thorough evaluation of generalization error for CBR prediction methods, including existing methods and those developed in this study. This involves a determination of how well the methods perform on new cases that were not used in constructing the model. For existing methods that are not based on the dataset assembled for this investigation, a simple check of model performance on this dataset can be used. However, for models developed here, more sophisticated validation techniques based on resampling of the full dataset are required to provide an accurate and honest estimate of generalization error.

1.4 Scope

Although the ultimate goal of the OLS effort is to use remotely-sensed information to estimate soil strength, it would be imprudent to begin model development with data that is collected in this manner. Finding a method that can perform the soil strength prediction task with data that has been directly measured from in situ testing should precede any such effort. Only a prediction method that has been built upon relationships among parameters measured with higher confidence, and tested to ensure acceptable performance with these types of inputs, can have any hope of succeeding when fed lower confidence input data—such as that which might be inferred from remote measurement. Therefore, this investigation will be limited to the use of high quality,

well-documented data that has been directly measured using standardized and consistent test methods.

In an attempt to address the wide variety of conditions under which soils can form and the broad ranges over which soil properties can be found in nature, this investigation has made every effort to collect and use data with substantial variety. In this manner, it is expected that the models developed with this dataset should be reasonably well-suited to generalizing to a wide variety of soils around the world for a broad range of naturally-occurring conditions.

It is possible to combine many of the various prediction methods that are available with another prediction method, optimization technique, or number system (such as fuzzy set theory). To keep the focus on choosing the best single method, without becoming preoccupied with all the possible hybridized approaches, this study was limited to evaluating prediction methods individually.

2 Background

2.1 California Bearing Ratio (CBR) Test

Since World War II, flexible and unsurfaced airfield pavement design has been primarily based on the California bearing ratio (CBR) design procedure, which relies on a companion soil strength index test. The determination of CBR is fundamental to assessing the type and amount of traffic loading that a pavement system can adequately support before it begins to fail. “The CBR test is determined by an arbitrary penetration procedure to obtain a modulus of shearing resistance of a subgrade or base course soil. This value is used to determine the required thicknesses of the various base courses through its application to empirically derived design curves” [14].

The California bearing ratio is a standardized test procedure recognized by both the American Society for Testing and Materials (ASTM) [15,16] and the United States Military [14,17]. The test is both a laboratory and a field test. It can be performed on laboratory compacted samples, on field samples carefully extracted from soil pits and returned to the laboratory, or on field samples in situ. The CBR test is not a direct measurement of a fundamental physical property, rather it provides an index from which strength can be assessed comparatively.

The CBR test is carried out by measuring the penetration resistance of a 1.954 inch diameter (3 square inch end area) cylindrical steel piston advanced into the soil at rate of 0.05 inches per minute. (Figure 2.1 shows the apparatus used to conduct the field in-place test.) The reaction force is measured, by means of a calibrated proving ring, at increments of 0.025 inches of penetration until a total penetration of 0.500 inch is reached. To determine the CBR index value, the reaction forces measured at 0.100 and

0.200 inch penetration are compared to standardized values of 1,000 psi and 1,500 psi respectively. These represent the resistance for a high-quality, well-graded crushed limestone gravel with ¾ inch maximum aggregate-sized particles. The two forces measured in the test, each divided by their corresponding standardized value and multiplied by 100 yields two index values, and the largest is reported as the CBR of the specimen in percent.

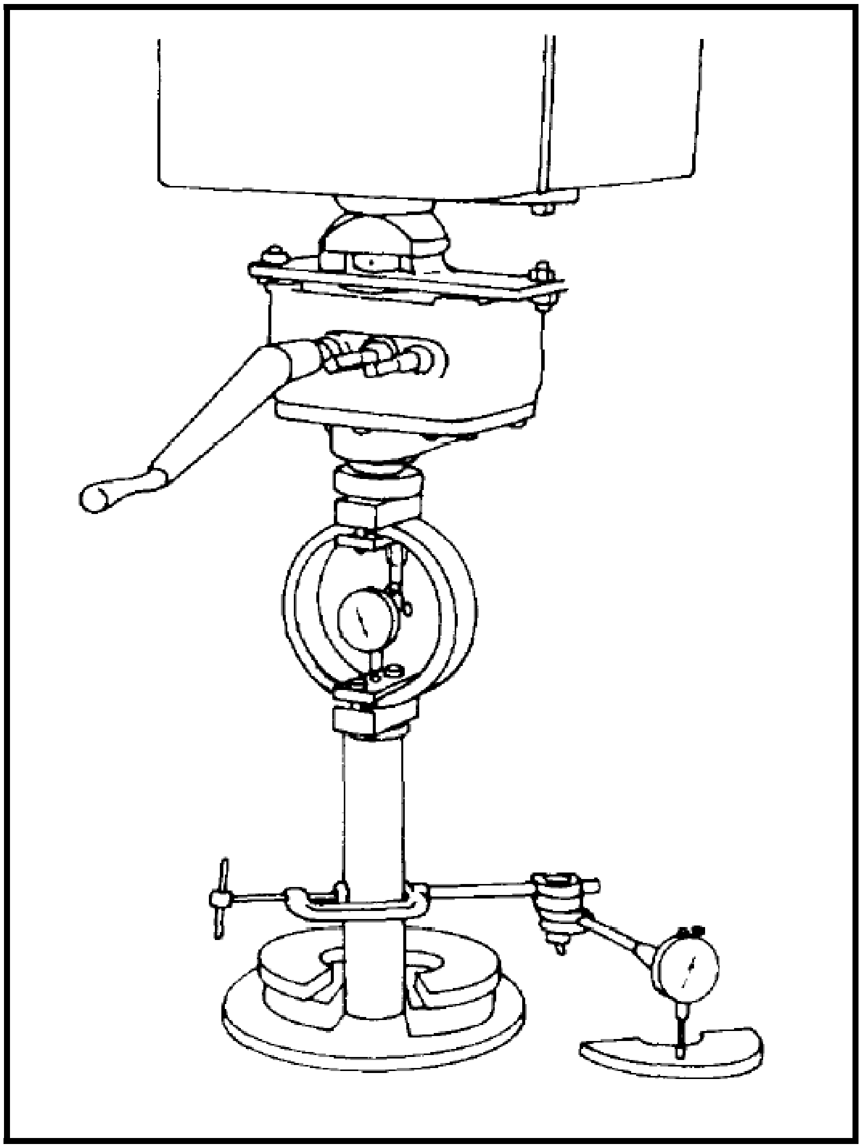

Figure 2.1: Apparatus for Field In-place California Bearing Ratio (CBR) Test [14].

When performing the test, measures must be taken in order to provide a confining pressure in the soil that is comparable to that which will exist during service conditions. In the laboratory, the specimens are compacted into a cylindrical steel mold that has an inner diameter of six inches. The lateral restraint provided by the mold is supplemented by annular surcharge weights through which the piston load is applied. These weights are chosen to be equivalent in mass per unit area to any layers above the material in the completed pavement design. Likewise, surcharge weights are used in the field test procedure to simulate the vertical confinement on the soil, however the surrounding soil provides the required lateral confinement pressure. The surcharge weights are the C-shaped pieces surrounding the lower end of the piston in Figure 2.1.

Caution in using the CBR test method for certain types of soils is advised. “The test is most appropriate and gives the most reliable results for fine-grained soils, although it is also used to characterize the strength of soil-aggregate mixtures (e.g., subbases) and unbound aggregate base courses. In cohesionless soils, especially ones that include large particles, the reproducibility of the test is poor” [18]. Because of this, the laboratory test procedure is limited to soils whose largest aggregate particle size is ¾ inch. In the case of soils where particle sizes greater than ¾ inch exist, these are removed from the sample and replaced by an equal mass of material that falls between the ¾ inch sieve and the number 4 sieve sizes (19 to 4.75 mm). “While traditionally this method of specimen preparation has been used to avoid the error inherent in testing materials containing large particles in the CBR test apparatus, the modified material may have significantly different strength properties than the original material. However, a large experience base has developed using this test method for materials for which the gradation has been

modified, and satisfactory design methods are in use based on the results of tests using this procedure" [15]. Obviously for field testing, the removal of larger-sized particles that may adversely affect the test results is not possible. Therefore, the in situ test is likely to encounter even greater problems in these types of soils than the laboratory measurement.

The California bearing ratio test was initially developed by O. James Porter for the California Highway Department during the late 1920s. "In time, Porter was able to develop the relationship between bearing ratios and pavement thicknesses for wheel loads up to 12,000 pounds and to correlate these curves with field performance" [19]. During World War II, when the military rapidly began fielding very heavy bombers and started to experience dramatic pavement failures, the Army Corps of Engineers extensively studied the CBR method for flexible pavement design and expanded it for use with much heavier loads. A board of expert consultants including Porter, Thomas A. Middlebrooks, George E. Bertram, and Dr. Arthur Casagrande of Harvard extended the CBR design curves up to wheel loads of 70,000 pounds based on a series of full-scale tests at a dozen airfields around the country.

CBR is criticized by some for being an empirical design method, however it is supported by more than 60 years of field experience under a wide range of conditions throughout the world. It still serves as the cornerstone of flexible and unsurfaced pavement design and evaluation today, especially for expedient and contingency evaluation of military airfield pavements [20]. Other design philosophies for flexible pavement do exist, including those with more of a basis in the theory of the mechanics of materials—such as layered elastic and finite element approaches. The rapid growth in

computing power and speed in the late 1980s made these methods feasible for use by practicing engineers and they have gained some acceptance [21].

Despite the advances in these state-of-the art approaches to pavement design, reliance on empirical observation remains central to pavement design. “Dependence on observed performance is necessary because theory alone has not proven sufficient to design pavements realistically” [22]. There remains a continued importance of and reliance upon the CBR method, in both military and civilian aviation. Strong statements in support of CBR versus mechanistic-empirical approaches have been made recently by leading pavement experts, including a senior consultant to the US Air Force and Boeing’s own chief pavements engineer:

> As I reflect on our trips to [Southwest] Asia and some of the questions we have dealt with from there these last few years, I have concluded that only the CBR method is robust and flexible enough to meet USAF needs for evaluating flexible airfield pavements in the theater of operations now or in the reasonable future….We need this CBR-type capability to assess totally non-conventional structures for aircraft operations and have a method with the reliability of CBR (even with its admitted imperfections). In 2000, I would have said it is time to retire CBR and USAF should shift to layered elastic analysis. These last years have proven me wrong and changed my mind completely….[U]ntil [the layered elastic method] can have robust, reliable criteria for all materials in the pavement structure (which to the extent we need, nobody I know of in the civil world is even talking about), it is not suitable for supporting USAF contingency and theater of operations requirements [23].

> In our view, the CBR Method has not been compromised for any other reason than being conservative, which is not such a bad idea, when one considers airfield pavements. On a related note, Mechanistic-Empirical [M-E] solutions are becoming more and more widely used every day, but do they accurately represent the effects of aircraft loading?…[W]e continue to find that the traditional tools render allowable stresses that are consistent with Modulus of Rupture values that are very familiar, but the results that we see when M-E tools are used are quite different, even when standard conditions apply [24].

Despite its empirical nature and the limitations of the test procedure, the CBR method clearly remains the most trusted and well-established criteria for determining the ability of a flexible or unsurfaced pavement structure to support the severe loads of aircraft traffic, especially for unconventional situations.

2.2 Existing CBR Prediction Methods

Some methods do exist to estimate the California bearing ratio of a soil, based on soil classification, other index test values, and/or physical property measurements of the soil. A few of these methods take a general approach and attempt to encompass many or all possible soil types, however most attempts have been limited in scope to a specific soil and only apply to one region, soil type, or specialized material.

2.2.1 "Universal" Approaches

2.2.1.1 Typical Values Based on the Unified Soil Classification System

The simplest approach to approximating the CBR value for a soil centers on typical values associated with soil classification. The Unified Soil Classification System (USCS) is a standardized technique for classifying soils for engineering purposes that is widely-used in the geotechnical community [25]. Within this system, soils are classified based on the distribution of their grain sizes and the cohesive properties of their fine material. The USCS grew out of the Airfield Classification System developed by Arthur Casagrande to train engineering officers at Harvard University during the Second World War [26]. In the USCS system, soils are divided into three major categories: coarse-grained materials, fine-grained materials, and highly organic soils. These categories are further divided into soil groups—the coarse-grained soils as either gravel

or sand and the fine-grained soils as either silt or clay. A letter symbol represents each of these four main soil groups, as shown in Table 2.1. These soil group letters are combined with a second letter, (shown in the lower half of Table 2.1) which is used to further describe the soil's characteristics. These descriptors include symbols to differentiate among grain size distribution, plasticity characteristics that describe cohesive behavior, and the nature of the organic material in a soil. For example, a sandy soil with few fines and a uniform grain size would be classified as a SP, or poorly-graded sand. A total of fifteen classes of these two letter combinations comprise the major soil types defined under the USCS system. Further designation for "borderline" soils are described by combinations of two of these fifteen major soil types. This occurs in cases where the fine material may be a combination of a clay and a silt (for example SC-SM designates a silty-clayey sand) or cases in which the amount of fines in a coarse-grained soil fall between 5 and 12 percent (for example GW-GM is a well-graded gravel with silt). Eleven of these combinations for borderline soils are generally recognized by the USCS system. It should be stressed to the reader that the USCS is a systematic and repeatable classification strictly based on test measurement values defined in the ASTM standard [25] and not a qualitative assessment of a soil based on subjective judgments. As such, the USCS class of a soil is inherently tied to the soil properties by which it is defined, and in the absence of these original classification test results can give some indication as to the range of grain sizes and plastic behavior that the soil is bounded by.

Table 2.1: Letter Symbols in the Unified Soil Classification System (after [14]).

Soil Groups	***Symbol***
Gravel	G
Sand	S
Silt	M
Clay	C
Soil Characteristics	***Symbol***
Well-graded	W
Poorly-graded	P
Low plasticity (liquid limit under 50)	L
High plasticity (liquid limit over 50)	H
Organic (silts and clays)	O
Organic (peat)	Pt

Guidelines for choosing CBR values based solely on USCS soil type are found throughout the literature. A summary of reported values from several of these sources is shown in Table 2.2. Generally, these are consistent for each soil type, with minor differences among the reported values. Part of this variation may be due to the fact that some refer to compacted soils [27], others refer to field-measured CBR values [14,28], while some do not specify test conditions [18,29–32].

Table 2.2: Typical California Bearing Ratio Values by Unified Soil Type.

USCS Soil Type	*USACE [30], US Army [31], and US Army & Air Force [27]*	*Yoder & Witczak [28]*	*US Army, Air Force & Navy [14] and PCA [29]*	*Rollings & Rollings [18]*	*NCHRP [32]*
GW	40 – 80	60 – 80	60 – 80	60 – 80	60 – 80
GP	30 – 60	35 – 60	25 – 60	35 – 60	35 – 60
GM	20 – 60	40 – 80	20 – 80	40 – 80	30 – 80
GC	20 – 40	20 – 40	20 – 40	20 – 40	15 – 40
SW	20 – 40	20 – 40	20 – 40	20 – 50	20 – 40
SP	10 – 40	15 – 25	10 – 25	10 – 25	15 – 30
SM	10 – 40	20 – 40	10 – 40	20 – 40	20 – 40
SC	5 – 20	10 – 20	10 – 20	10 – 20	10 – 20
ML	15 or less	5 – 15	5 – 15	5 – 15	8 – 16
CL	15 or less	5 – 15	5 – 15	5 – 15	5 – 15
OL	5 or less	4 – 8	4 – 8	4 – 8	--
MH	10 or less	4 – 8	4 – 8	4 – 8	2 – 8
CH	15 or less	3 – 5	3 – 5	3 – 5	1 – 5
OH	5 or less	3 – 5	3 – 5	3 – 5	--
Pt	--	--	--	< 1	--
CL-ML	--	--	--	--	--
GW-GM	--	--	--	--	35 – 70
GW-GC	--	--	--	--	20 – 60
GP-GM	--	--	--	--	25 – 60
GP-GC	--	--	--	--	20 – 50
GC-GM	--	--	--	--	--
SW-SM	--	--	--	--	15 – 30
SW-SC	--	--	--	--	10 – 25
SP-SM	--	--	--	--	15 – 30
SP-SC	--	--	--	--	10 – 25
SC-SM	--	--	--	--	--

2.2.1.2 Mechanistic-Empirical Design Guide

Another general approach to the problem of estimating CBR has been developed as a part of the highway pavement community's recently released *Mechanistic-Empirical Design Guide for New and Rehabilitated Pavement Structures* [32]. The design guide methodology includes three tiers of confidence in the resulting pavement designs,

depending on the quality of input data provided to the model. This ranges from the highest level, where the design is based on a detailed, project-specific series of laboratory characterization tests on the construction materials, to the lowest level where default values based on simple material characterization tests and/or regional norms are used as model inputs. One of the parameters needed to perform a flexible pavement design using this system is the resilient modulus, which is "a specific type of modulus of elasticity that is based on the recoverable[†] strain instead of total strain" [18]. In order to provide an estimate of the resilient modulus parameter for the lower tiers of the system where it is not measured directly, an appendix to the Mechanistic-Empirical Design Guide was developed that relates resilient modulus to much simpler soil characterization tests by way of CBR as an intermediary step.

The correlation between soil index properties and California bearing ratio for this method is based on a simple regression approach. Separate relationships were determined for coarse-grained soils that exhibit no cohesive behavior (GW, GP, SW, and SP) and for soils with more than 12 percent fines that exhibit plastic behavior (GM, GC, SM, SC, ML, MH, CL, and CH). The CBR values were selected by choosing average values for each USCS soil type based upon sources that provide typical CBR values by classification, as illustrated in the previous section. The index property values were selected by examining the USCS classification criteria for each soil type, and choosing a typical value for that USCS soil type. The index properties chosen to correlate with CBR included:

[†] When a soil is loaded and unloaded cyclically it accumulates some amount of permanent plastic strain. The "recoverable strain" only includes the average deformation between each of the loaded and unloaded states, and not the permanent plastic strain.

D_{60} Diameter on the cumulative size distribution curve where 60 percent of particles are finer (in millimeters)

P_{200} Percent passing (finer than) the number 200 sieve size (in decimal form)

PI Plasticity index (in percent)

The last two properties were combined into a composite index called the *weighted plasticity index*. This term, denoted by wPI, is defined in the method by:

$$wPI = P_{200} \bullet PI \qquad (if \quad PI > 0) \tag{2.1}$$

For the clean, coarse-grained, non-plastic soils where $wPI = 0$, the CBR were correlated with D_{60}. The method provides the following prediction relationship for these:

$$CBR = \begin{cases} 5 & (\ if\ D_{60} \leq 0.01mm\ ) \\ 28.09\left(D_{60}\right)^{0.358} & (\ if\ 0.01mm < D_{60} < 30mm\ ) \\ 95 & (\ if\ D_{60} \geq 30mm\ ) \end{cases} \tag{2.2}$$

For the second group of soils that exhibit plastic behavior, a different correlation for CBR was determined. In cases where the soil has a fines content (P_{200}) greater than twelve percent and the weighted plasticity index (wPI) is nonzero, the prediction equation is:

$$CBR = \frac{75}{1 + 0.728\left(wPI\right)} \tag{2.3}$$

The R^2 value on the training set data for these two equations were reported as 0.84 for the coarse-grained materials and 0.67 for the plastic materials.

2.2.1.3 Soil Strength "Signature" Concept

Another generalized CBR prediction technique grew out of efforts to develop an analytical model to assess the impact of soil-moisture variations on unpaved road performance [33]. Most pavement design procedures characterize the strength of each layer in terms of a single, generally conservative value that represents the worst-case, fully-saturated state of the material. For California bearing ratio, this involves testing lab-prepared samples soaked for four days so they reach a moisture condition close to full saturation—commonly referred to a "soaked CBR." However, most of the time the circumstances in the pavement layers are actually much more favorable and the soil can carry higher loads without damage.

In order to capture the effect that higher in situ strengths could have on accommodating greater traffic volumes, researchers developed an integrated performance prediction model including an element that correlates CBR to soil type, moisture content, and density. A literature search gathered fifty-three sets of test results for twenty-one USCS classification groups representing materials ranging from high plasticity clays to well-graded gravels and sands. These datasets are typical of the type of laboratory test suites that are usually performed to characterize the relationships between moisture & density, density & strength, and moisture & strength. These include moisture-density curves (known as Proctor curves) and moisture-CBR curves, both over a range of different compaction energies. The method developed here in effect generalized these relationships for each soil type, and then proceeded to combine these two test results into a single plot. In this way a series of curves, representing different levels of compaction (i.e., density), were drawn in a CBR-moisture content space for a single USCS soil type.

This allows the prediction of a CBR value based on known moisture content and density for each USCS soil type.

One example of this method, termed a "soil strength signature," for a silty-clayey sand (SC-SM) is provided in the report describing the integrated performance prediction model [33]. This plot is shown in Figure 2.2. Relationships for the other USCS soil types exist in another report [34], but this literature source could not be located. Therefore, no subsequent testing to determine how well the method worked on the dataset collected for this study could be carried out.

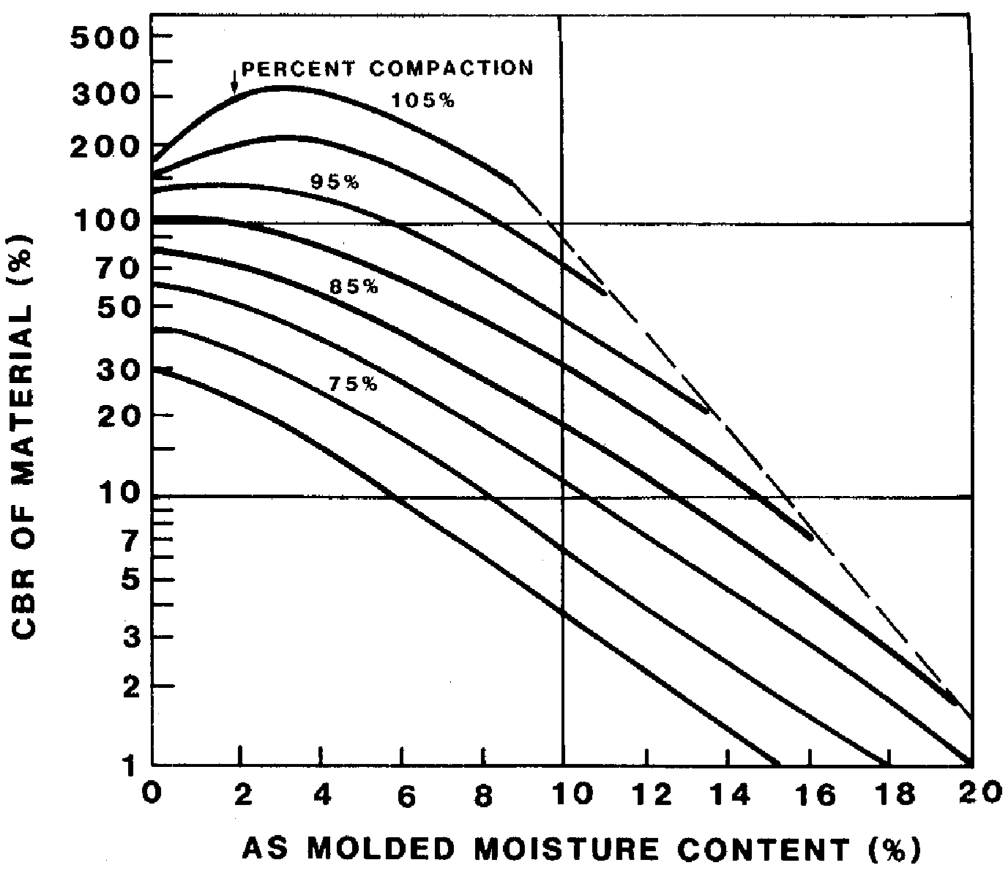

Figure 2.2: Soil Strength Signature Showing the Relationship Among Moisture Content, Density, and Strength for an SC-SM Material (from [33], with permission from ASCE).

2.2.1.4 Joint Rapid Airfield Construction (JRAC)

One more generalized approach to predicting CBR values involves the Joint Rapid Airfield Construction program. The goal of the JRAC effort is to enable a rapid assessment of a soil with a miniaturized field soil laboratory kit, so that critical construction parameters such as USCS soil type, compaction curves, and design CBR values can be estimated within a one hour timeframe. Though the study is still in progress and the method has not yet been published, the work is based on correlating CBR to moisture content for different USCS soil types and compaction levels. Based on a worldwide dataset of soils from both US and overseas locations [35,36], the JRAC approach uses a regression analysis. The CBR is correlated with a term called the "normalized moisture content" through a fourth order polynomial curve-fit. The normalized moisture content is defined as the natural gravimetric moisture content of the soil minus the optimum moisture content of the soil for a standardized compaction energy as determined by a Proctor test. Limits are placed on how far above and below the optimum moisture content the relationship is considered valid, to prevent extrapolation into areas without supporting data or where the curve-fit is poor.

2.2.2 Specific and Specialized Approaches

2.2.2.1 Routine Laboratory Tests for Site Characterization

In order to determine the engineering properties of a soil (or soils) on a specific construction site, a set of representative samples are obtained and a series of laboratory tests are performed. Usually this suite of tests include a grain size analysis, liquid & plastic limit determination, moisture-density (Proctor) tests at different levels of

compaction, and specific gravity. They also tend to include sets of California bearing ratio tests on samples made over a range of molding moisture contents at several levels of compaction energy, in both soaked and unsoaked conditions. This type of information is routinely used for design and construction specifications. While not a formal method of prediction for field CBR, these basic index property tests permit a classification of the soil and provide some indication of how these properties can affect the soil strength. A typical explanation of the purpose behind this approach appears in many of the pavement evaluation reports from which much of the data for this study was collected: "Comparison of the laboratory data with in situ data provides an insight into the behavior of soils under differing conditions and brings to light possible explanations for poor performance of certain soils" [37].

Obviously, these tests are intended to provide information for particular soils at specific geographic locations. However, they can form the basis for more wide-ranging CBR prediction methods. For example, the soil strength signature concept described above is actually a formal attempt at generalizing the relationships among this suite of tests to allow more widespread predictions to be made. And the datasets for the JRAC work and for this study are mainly comprised of these test results from many different locations around the world, allowing for some validity in a global application of the techniques.

2.2.2.2 Relationships for Regional Soils

Several attempts at predicting the California bearing ratio of soils for a specific soil or geographic location can be found in the literature. Three papers by Attoh-Okine describe the application of different prediction methods to the problem of estimating the

strength of lateritic soils, a material which forms under the unique conditions found in the humid tropics [38–40]. The utility of laterites for paving applications can be tricky to assess, since the specifications commonly used for more conventional soils in the temperate regions do not seem to be valid for this unusual soil. Therefore, common practice for lateritic materials has focused on the direct measurement of strength as an acceptance criteria for construction applications. In order to draw a direct correlation between strength and some common index tests, the first paper begins with a multivariate regression approach. The dataset for this analysis consisted of thirty-eight cases, collected along the Trans-African Highway in the rainforests of southwestern Ghana. The features in the dataset included field dry density, maximum dry density, relative compaction, field moisture content, optimum moisture content, field CBR, laboratory CBR (soaked), liquid limit, and plasticity index. Based on a sensitivity analysis of these features, the field dry density, plasticity index, and liquid limit were selected to include in a linear multiple regression. The recommended model was reported as:

$$FCBR = 89.676 \cdot FDD + 7.163 \cdot PI - 7.65 \cdot LL \tag{2.4}$$

where:

$FCBR$	=	Field CBR (in percent)
FDD	=	Field dry density (in grams per cubic centimeter)
PI	=	Plasticity index (in percent)
LL	=	Liquid limit (in percent)

The R^2 value on the training set data was reported as 0.94 and the standard error as 22.18. No estimation of a generalization error rate on an independent test set was performed.

Attempts to apply artificial neural network analysis[‡] to a slightly larger dataset (45 cases) were also reported [39,40]. Several types of networks, including those based on feed-forward back-propagation, generalized adaptive techniques, and genetic algorithm learning, were evaluated. For the feed-forward back-propagation (FFBP) neural network, the same input features from the regression analysis (with the exception of soaked laboratory CBR) were used to predict field CBR. The network architecture was fully connected with seven input nodes, two hidden layers (five neurons and two neurons respectively), and one output neuron for the CBR target. Based on network training with a 55% to 45% ratio of training to test set cases, the correlation (R) between the neural net prediction and the known target values was reported as 0.580 for all data. Subsequent trials with an adaptive network architecture, where connections and nodes are added sequentially, resulted in an improved performance with $R = 0.735$. In this case, only five of the seven input features were selected by the adaptive model, with maximum dry density and relative compaction being eliminated. In a final attempt to improve the performance of the neural network, a single hidden layer architecture with a genetic algorithm learning method was used. In this case the root mean square error on the test set ranged from 18 to 19.7, depending on the number of inputs ranging from one to four. The study found that the neuro-genetic algorithm was the most efficient of the three and better able to identify appropriate input data. "The analysis shows the most important parameter needed for modeling of lateritic soil strength in terms of CBR is field [dry] density. Although other parameters like field moisture content and Atterberg limits appear to be very important, quantitatively they can be eliminated in most modeling of strength characteristics of lateritic soils" [40].

[‡] A detailed discussion of artificial neural networks can be found in section 4.2.3.

Some issues with the analyses of the lateritic gravels and the reported performance are apparent. In the case of the FFBP network, the architecture of the neural net contains 55 weights and biases that need to be optimized. However, with only 25 training cases, it is obvious that the problem is overspecified. Subsequent trials with the adaptive network architecture and the genetic algorithms produced significantly less complex models, which are more appropriate to the number of available training cases. In addition, the goodness-of-fit and error metrics reported for all of the prediction techniques are different and (unfortunately) do not permit a direct comparison of performance among the various methods to be made independently.

Other reports of CBR prediction for regionally specialized soil types include estimates for subgrade soils common to Kuwait and undisturbed clay soils found in Britain. In the case of the Kuwaiti soils, optimum moisture content and maximum dry density for the desired level of compaction were found to be good predictors of California bearing ratio [41]. For the British clays, the CBR was estimated by idealizing the test as a bearing capacity failure problem [42]. From this approach, the bearing capacity is inferred by way of a soil's suction and angle of friction, which in turn are estimated from the liquid limit, plastic limit, and moisture content of the soil. The resulting predictions for CBR as a function of moisture content appeared to be accurate at higher densities, but had problems with low density soils.

2.2.2.3 Specialized Materials

In some cases, the utilization of specialized recycled materials in pavement construction has resulted in the establishment of CBR correlations for these non-traditional sources of material in order to provide guidance and acceptance criteria

for their use. One such instance involves sand discarded after casting processes in the foundry industry, known as excess system sands, being used for highway applications [43]. The material is typically composed of uniformly graded silica sand, with added clay binders ranging from 5 to 15 percent by weight in order to provide enough cohesive behavior so that molds can be easily formed for casting the molten metal. A series of index tests, including CBR, were performed on a variety of these foundry sands and a multivariate stepwise linear regression carried out with the results. The analysis encountered difficulty in trying to find a regression model that could be used for both plastic and non-plastic soils, so two separate relationships were reported. For non-plastic excess system sands the relationship was reported as:

$$CBR = 32.4\gamma_{dm} - 1.93P_{200} - 2.64R_o - 361 \tag{2.5}$$

where:

CBR	=	California bearing ratio (in percent)
γ_{dm}	=	Standard Proctor maximum dry unit weight (kN/m^3)
P_{200}	=	Percent passing (finer than) the number 200 sieve size
R_o	=	Krumbein [44] particle roundness (in decimal form)

For plastic excess system sands, the relationship was reported as:

$$CBR = 178R_o - 7.6\gamma_{dm} - 4.25AC \tag{2.6}$$

where:

AC	=	Active clay content (in percent)

and the other factors are the same as the previous equation

The R^2 value on the training set data were reported as 0.94 for the non-plastic correlation and 0.86 for the plastic. The authors provide a physical explanation for the non-plastic

relationship by asserting that “Shearing resistance of cohesionless soil is strongly influenced by interparticle friction, which increases with higher dry unit weight. An increase of fines or particle roundness causes a decrease in interparticle friction and lower CBR” [43]. The physical meaning behind the plastic correlation equation is not clear, so no explanation is given.

3 Data

Because of the diverse, demanding, and time-sensitive nature of military operations, decision support systems—such as those being developed under the Opportune Landing Site program— must be applicable to the broadest possible range of locations and conditions that are likely to be encountered. In order to fulfill this objective, special consideration and attention were taken in acquiring the dataset used for the analysis and modeling in this investigation.

From the beginning, it was apparent that the dataset would need to meet several unique requirements to be suitable for generating useful relationships. The constraints that guided the search for data included the following objectives and reasoning:

1. Attempt to incorporate as many of the 26 USCS soil types into the database as possible. Because they are based on separating different regimes of engineering behavior in soil, a diversity of USCS classes should expose machine learning methods to all the mechanisms that drive soil strength.

2. Ensure that the database is representative of the relative prevalence of the USCS soil types worldwide. In effect, the data should reflect to some degree how likely are we to encounter each of the different soil types in practice and encompass the larger variety that can be present in some of the more common soil types.

3. Focus specifically on geotechnical parameters, especially those typically used to characterize engineering behavior in the civil engineering community.

4. Concentrate on records that contain actual California bearing ratio measurements, not another soil strength index or parameter that can only be correlated to CBR.

5. Make sure that the data encompasses the range of conditions that we would expect to find in naturally-deposited soils. In this respect, care must be taken to ensure that the acceptance criteria placed on construction materials do not skew the database. For example, standardized laboratory tests limited to high quality material could reflect higher densities, lower fines contents, and lower moisture contents.

6. Incorporate as much geographic, geologic, environmental, and depositional diversity as possible. In this manner, there is some attempt at trying to reflect the wide variety of unique conditions under which natural soils can form.

7. Bring together a consistent and well-documented dataset. The use of standardized test methods is critical for high confidence. Ensuring that individual data records are tied to their original sources can be useful in many respects: any peculiar soils could be isolated and dealt with separately if necessary, further information may be collected from documented sources to support future efforts, and inferences due to tests locations or seasonal variation might be possible.

These principles formed the basis for evaluating prospective sources of data for the OLS soil strength prediction study and the design of the database.

3.1 Literature Search

As part of a thorough survey, many different sources of data were considered as possible candidates for compiling the OLS CBR Database. These include technical reports containing detailed geotechnical test results, soil mapping and soil survey efforts from the soil science and agricultural communities, airfield pavement evaluation reports generated by USACE and USAF to monitor and assess these facilities, collaboration with parallel research efforts within the Corps of Engineers, and finally some emerging online and commercial geotechnical databases. Some of these sources proved to be incompatible with the constraints and objectives outlined in the previous section. In some cases, however, the sources that could not be utilized for the OLS CBR Database did prove useful in other ways. A few of the resources proved useful in developing a schema for this effort. And some of the efforts underway to develop geotechnical databases should provide much better opportunities for data mining and machine learning approaches in the future when they are complete.

Some of the soil mapping and soil survey work that was considered included global efforts at cataloging the world's soil resources. The United Nations Food and Agriculture Organization (FAO) produced a world soils map in the 1970s [45]. A more recent effort is underway to update this map into an electronic Soil and Terrain Database (SOTER) product at much finer scale than the original [46]. Unfortunately, since the focus for these maps was agricultural productivity of the soils, there was very little specific engineering data that could be gleaned from them. The gross scales of these mappings, ranging from 1:5 million for the earlier map down to 1:250,000 for the SOTER effort, are inadequate for the OLS objectives. In addition, the system of taxonomy used to describe soils in these maps are qualitative and our ability to correlate these directly with the USCS system is tenuous at best. Despite these shortcomings, the SOTER methodology for classifying landforms, lithology of soil parent material, depositional processes, and clay mineralogy [47] were found to be very helpful and they were adopted for use in the OLS CBR Database schema.

Parallel research efforts within the US Army Corps of Engineers were also consulted for use in fabricating the OLS CBR Database. The soils database compiled for the Joint Rapid Airfield Construction program, described earlier in section 2.2.1, was evaluated for use in this investigation. Unfortunately, the data collected for this work focused on providing a general summary of soil parameters and not the specific input-output pattern cases that are required to train machine learning algorithms. Other current research at the ERDC is focused on soil strength from a ground vehicle mobility perspective that concentrates on the cone index (CI), a soil strength index test based on the static penetration of a 30° cone that overlaps the lower end of the CBR range [48,49].

The Fast All-season Soil STrength (FASST) model developed to predict the state of the ground in the theater of operations includes the ability to forecast this soil strength index based on soil type and changing weather conditions [50]. However, the basis for the soil strength calculations is a model that relies only on a single exponential correlation between CI and moisture content for each USCS soil class [51]. A twin program under the OLS project to collect a database of CI related measurements is also in progress. Because these vehicle mobility database efforts do not focus on tests containing California bearing ratio measurements, they were not directly useful for the CBR prediction task.

Another body of soil data considered for the OLS CBR Database included some existing and emerging electronic geotechnical databases. A commercial off-the-shelf relational database containing six thousand distinct soils called *SoilVision* was evaluated [52]. Even though the database is well organized and has fields for many of the engineering parameters we wanted to incorporate in the OLS CBR Database, the existing dataset included in this package concentrated mostly on hydraulic properties of soils and had little CBR information. Another existing soil database maintained by the Natural Resources Conservation Service (formerly the National Soil Conservation Service) contains some textural, plasticity, grain-size uniformity, density, and moisture content test data, however it is focused on agricultural use and lacks any strength data that is critical for the current analysis [53]. Efforts are underway by the National Geospatial-Intelligence Agency to build a global soils database [54] by digitizing unpublished USDA 1:1 million soil maps, but as with the SOTER mapping initiative the scale and focus are not immediately useful for the OLS task. Another more relevant

initiative is underway by the US Air Force called GeoBase that aims to collect and archive data related to their bases worldwide [55]. Included in this database will be information on pavement and soil data gathered in conjunction with construction projects, condition assessments, and airfield pavement evaluation report generation. While current activities do not collect CBR information directly, this dataset may prove useful to other data mining efforts when it becomes available. Incorporation of historical test data into this framework would also be valuable, especially for the OLS program. A final resource that may allow greater accessibility to geotechnical data is Geotechnical Markup Language, an open source hypertext markup language scheme for soil data with an engineering focus [56]. If this initiative catches on, then future data miners could use this online international repository to search for new correlations.

Ultimately, the most valuable resources turned out to be the technical reports, and the airfield pavement evaluation reports. These contain a wealth of in situ field test and laboratory characterization data for a wide variety of soils from around the United States and locations around the world where DoD currently maintains bases or has in the past. Two technical reports were selected for use in the OLS CBR Database. The first details an early study carried out by the Army Corps of Engineers immediately following World War II, which investigated moisture conditions under flexible airfield pavements [57]. Eleven field locations around the continental United States served as the test locations. The airfields chosen were located in arid, semiarid, and humid regions with minimal frost exposure. Previous attempts to measure moisture content with sensors proved unsuccessful with the technology available at the time. Therefore, measurements of soil properties, including numerous field CBR tests, were made directly in soil pits and

boreholes dug within the pavement sections and adjacent non-paved areas. The availability of these field readings coupled with thorough laboratory characterization tests performed on the same materials made the report a particularly valuable repository of data relevant to the current investigation. A second technical report, involving a recent round of full-scale tests to help certify the C-17 airframe for unsurfaced airfield operations, was also used [58]. This report contains detailed field test data from six semi–prepared runway locations mainly in the southwest United States. The final resource used in the database included many airfield pavement evaluation reports. These documents are produced for Army, Air Force, and Navy facilities on a regular basis to monitor pavement conditions over time, certify them for operational use by different aircraft, and to help in planning ongoing maintenance and new construction projects. These reports contain extensive field and lab test results used in this process that tend to be very consistent, due to the well-documented standard test methods they were based on [14]. However, due to the shift in testing procedures towards non-destructive techniques in the 1990s, earlier evaluations that relied on excavation of subsurface test pits below the pavement proved to be the most valuable. Because the destructive tests involve significant time, expense, and disruption of operations, they are very rarely carried out today. This makes this historical dataset a particularly unique resource that should be carefully preserved.

3.2 Compiling CBR Database from the Literature

Data collection for the OLS CBR Database took place in two phases, each yielding approximately half of the cases in the final dataset. The first phase focused on

the two Army Corps of Engineer technical reports discussed above [57,58]. A second phase concentrated on the airfield pavement evaluation reports.

A considerable number of pavement evaluations were available and they needed to be prioritized in terms of their value for the OLS CBR Database. In a hard-copy archive at ERDC-CRREL containing evaluations from the 1940s to the present, an estimated 871 reports were catalogued (Figure 3.1). A second archive, kept by the Air Force Civil Engineering Support Agency (AFCESA), was surveyed during March 2005 [59]. This repository contained an undetermined number of evaluations from the 1960s onward, which were scanned into electronic format. Working with the electronic archive due to ease of access and sharing, reports containing test pits with CBR field test measurements were catalogued. A total of 937 pits from 161 airfield pavement evaluation reports were identified. For each report the number of pits containing CBR information and a general ranking of the USCS soil types present were recorded. Using this information, a prioritization scoring system was created for these reports to estimate the amount of useful data in each and guide the data entry process. A composite score was assigned to each, incorporating the number of CBR pits in a report, the relative prevalence of each soil type for that site, and the degree of need for that soil type in the database after the first phase. In this way the evaluation reports were ordered so the highest-ranked might provide the most data for the soil classes that were still lacking in the database. Because of the unique complexity of organic soils and the lack of these soils in constructed airfields due to their undesirable engineering properties, organic soils were not deliberately targeted for collection. Ultimately, a total of thirty-two airfield pavement evaluation reports representing seventeen locations within the continental

United States (CONUS) and twelve bases outside the continental US (OCONUS) were entered into the OLS CBR Database [37,60–90].

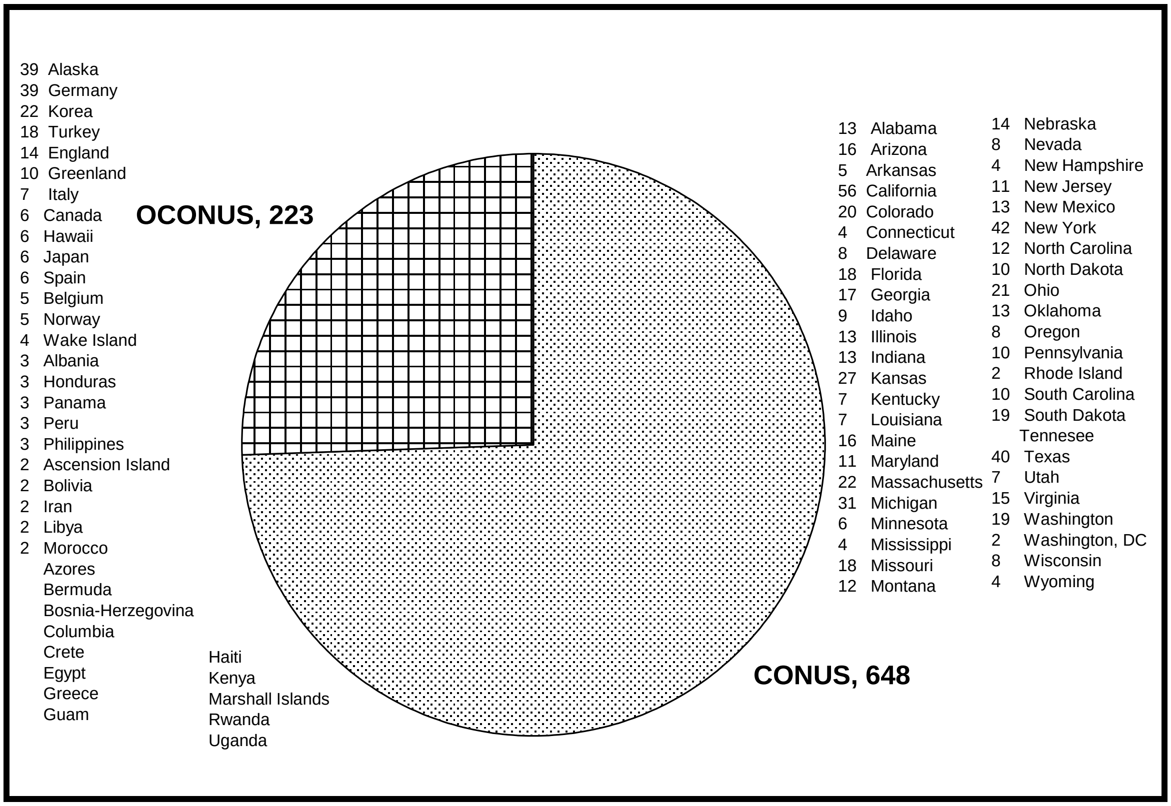

Figure 3.1: Number of Airfield Pavement Evaluation Reports in ERDC-CRREL Archive by Location.

From the technical and airfield pavement evaluation reports, the information detailed in Table 3.1 was compiled for use in training and evaluating models. A total of sixty-two fields were chosen to store information about data identification, reference source documentation, sample site description, soil classification, physical property data, strength index testing (both laboratory and field), particle sizes and shapes, and remarks. The contents for each of these fields in described in further detail in Appendix C. Features were chosen by consulting with a group of subject matter experts to determine a broad range of data types that may either have a qualitative relationship to soil strength or allow inferences to be made about soil conditions. Even though many were not filled

either at all or to a significant degree, this large number of fields were useful in providing a comprehensive scheme for all data types that might be found in any of the literature sources, flexibility for future data collection, and crossover with other databases (such as the OLS cone index work) for possible merging at a later date. The manner in which grain size data was presented in the airfield pavement evaluation reports necessitated two fields for each particle size. In many of these reports, similar soils were grouped together into families and a band was used in the plot of grain size distribution. A maximum and a minimum value were used to capture this range of particle sizes for each soil, with the intent that this might be useful for methods that could handle probability distributions.

The OLS CBR Database required considerable effort to assemble. The use of optical character recognition software to capture the data from the documents was explored, however the table and graph formats in the reports did not lend themselves to this technique. Ultimately, manual data entry was used which proved to be slow, however this deliberate approach did provide some benefits. The methodical approach yielded a consistent dataset and error checking provided a high degree of data integrity that allowed confidence during subsequent analysis work.

Table 3.1: Fields* in Opportune Landing Site California Bearing Ratio Database.

OLS Data Point #	Moisture Content as Tested (weight %)
JRAC Soil #	Moisture Content as Tested (volumetric %)
Test or Sample Date	Trafficability Cone Index (CI)
Report #	Remolding Index
Report Date	DCP Index (dynamic cone penetrometer)
Report Title	Field CBR
Country Code (ISO-3166§)	Field Dry Density
Location	Field Wet Density
Test Station	¾ inch Sieve, Maximum Percent Passing
Layer	¾ inch Sieve, Minimum Percent Passing
Landform	⅜ inch Sieve, Maximum Percent Passing
Lithology of Parent Material	⅜ inch Sieve, Minimum Percent Passing
Deposition Type	#4 Sieve, Maximum Percent Passing
Depth to Water Table	#4 Sieve, Minimum Percent Passing
Soil Type, USCS	#10 Sieve, Maximum Percent Passing
Alternate Soil Type	#10 Sieve, Minimum Percent Passing
Alternate Soil System	#40 Sieve, Maximum Percent Passing
Soil Description	#40 Sieve, Minimum Percent Passing
Clay Mineralogy	#100 Sieve, Maximum Percent Passing
Specific Gravity	#100 Sieve, Minimum Percent Passing
Sample Depth Below Grade	#200 Sieve, Maximum Percent Passing
Plastic or Non-Plastic	#200 Sieve, Minimum Percent Passing
LL (liquid limit)	0.005 mm, Maximum Percent Passing
PL (plastic limit)	0.005 mm, Minimum Percent Passing
PI (plasticity index)	0.001 mm, Maximum Percent Passing
Compactive Effort	0.001 mm, Minimum Percent Passing
Molding Moisture Content	Roundness, Gravel
Dry Density (laboratory)	Roundness, Sand
Optimum Moisture Content and Max. Density	Sphericity, Gravel
Unsoaked CBR (laboratory)	Sphericity, Sand
Soaked CBR (laboratory)	Remarks

* See Appendix C for a detailed description of each field.
§ Two letter standard code from the International Standards Organization [91].

3.3 Geographic and Soil Type Distribution

A total of 4,592 cases[§] of separate field test conditions were collected from all sources. Approximately one third of these (1,580) contained information regarding the California bearing ratio. The remaining two thirds were collected because it was easier to record all the data from each report during the data entry process. Also, these cases provided useful soil condition information for determining relationships among the non-CBR features and could be valuable in further data mining efforts not focused on CBR. For 47 cases, non-numeric CBR data were recorded (e.g., CBR $\geq$ 100) in case they could be used for classification or probability distribution based models. However, most of the CBR cases (1,533) had a numerical value for the strength index, and these are the ones that were utilized.

The data collected for the Opportune Landing Site California bearing ratio Database came from forty-six separate test sites, shown in Table 3.2. The number of cases are listed for each site, both for the full dataset and a subset of only those containing the numerical CBR target value. These sites include thirty-four from within the continental US, seven located around the Pacific Ocean, and five from in or near Europe. The geographical distribution of these sites, shown in Figure 3.2, Figure 3.3, and Figure 3.4, represent a good variety of different locations around the world. They encompass a broad range of geologic and environmental conditions, such as arid deserts, humid tropics, glacial till, coral islands, alluvial plains, volcanic deposits, dry lakebeds, and frost-active areas. Therefore, they should contain many of the different combinations of conditions and processes that lead to soil formation.

[§] Originally, 4,608 cases were collected in the database, but sixteen were eliminated before further analysis because they were either stabilized with cement (10) or had compaction energy of CE 26 that differed (6) from all other cases which were CE 55 [17].

A summary of the Unified Soil Classification types contained in both the full database and the numerical CBR subset appears in Table 3.3. These give some indication of the variety of soils included in the entire dataset and in the model training set. In order to get some sense of how well the database represented global soils, a comparison was made to an existing estimate of worldwide prevalence of USCS soil types [92]. Figure 3.5 shows the percentage distribution of each soil type relative to the total number in each dataset, while the associated values from the literature are an estimated percentage based on overall land area. The chart shows that the distribution in the numerical CBR subset tracked the overall database quite closely. Some exceptions to this include a slight increase in the number of gravel soils (GW, GP, GM, GC) and a significant decline in low plasticity clays (CL) and high plasticity silts (MH) for the CBR cases.

The differences in the database distribution and the worldwide estimate are more significant, but some similarities do exist. SM, CL, CH, and SC soils are respectively the most common soils in the worldwide estimate, while SM and CL— followed by SC— are most common in the CBR subset. The CL and CH soils are slightly under represented in the database, and the dominance of SM soils over all others is not present. Also, the worldwide estimate contains no gravel soils (USCS classes beginning with a G), while these are quite common for the database. This reflects the fact the data collection concentrated on airfield pavement structures, which are designed with granular base and subbase material. The number of ML and SP soils are greater in the database than would be expected from the worldwide estimate. The reasons for this are not entirely clear, but we speculate that these soil types may be most common on smooth, flat landforms where airfields are likely to be placed. Also, very few CL-ML soils are found

in the database compared to the worldwide estimate, and, by design, organic soils (OL, OH, and Pt) were specifically not included in the database. While it is unclear whether either the database or the estimate from the literature represent an accurate assessment of the worldwide distribution, the database clearly exhibits a reasonable diversity of USCS soil types.

Table 3.2: Number of Cases in CBR Database and Subset by Test Location.

Airfield Name	Location	ICAO Code	Total Cases	CBR* Cases
Alamo Landing Zone	Alamo, Nevada	KL92	12	12
Andersen Air Force Base	Yigo, Guam	PGUA	56	31
Bergstrom Air Force Base	Austin, Texas	KAUS	280	100
Bicycle Lake Army Airfield	Fort Irwin/Barstow, California	KBYS	13	13
Cannon Air Force Base (formerly Clovis Air Force Base)	Clovis, New Mexico	KCVS	231	45
Castle Air Force Base	Atwater, California	KMER	80	16
Clark Air Base	Angeles City, Philippines	RPMK	103	33
Craig Air Force Base	Selma, Alabama	KSEM	105	65
Creech Air Force Base (formerly Indian Springs Airfield)	Indian Springs, Nevada	KINS	61	19
Edwards Air Force Base (Rogers Dry Lakebed)	Edwards, California	KEDW	5	5
Eglin Air Force Base	Valparaiso, Florida	KVPS	92	23
Eielson Air Force Base	Fairbanks, Alaska	PAEI	71	21
George Air Force Base	Victorville, California	GAFB	190	46
Goodfellow Air Force Base	San Angelo, Texas	KGOF	189	85
Hancock Field Air National Guard Base	Syracuse, New York	KSYR	38	17
Hickam Air Force Base	Honolulu, Hawaii	PHNL	126	54
Holland Landing Zone	Fort Bragg, North Carolina	--	1	0
Holloman Air Force Base	Alamogordo, New Mexico	KHMN	163	39
Kadena Air Base	Okinawa, Japan	RODN	277	55
Keesler Air Force Base	Biloxi, Mississippi	KBIX	105	45
Kingsley Field Air National Guard Base	Klamath Falls, Oregon	KLMT	140	30
Kirtland Air Force Base	Albuquerque, New Mexico	KABQ	294	94
Loring Air Force Base	Limestone, Maine	KLIZ	67	33
Luke Air Force Base	Glendale, Arizona	KLUF	51	28
Marana Air Park	Marana, Arizona	KMZJ	122	33
Maxwell Air Force Base	Montgomery, Alabama	KMXF	78	12
McChord Air Force Base	Tacoma, Washington	KTCM	41	25
McGuire Air Force Base	Wrightstown, New Jersey	KWRI	117	29
Memphis International Airport (formerly Memphis Municipal Airport)	Memphis, Tennessee	KMEM	147	71
Myrtle Beach Air Force Base	Myrtle Beach, South Carolina	KMYR	108	31
Nellis Air Force Base	Las Vegas, Nevada	KLSV	107	20
Quonset State Airport	North Kingstown, Rhode Island	KOQU	60	21

Airfield Name	*Location*	*ICAO Code*	*Total Cases*	*CBR* Cases*
RAF Mildenhall Air Base	Suffolk, England	EGUN	57	16
Reese Air Force Base (formerly South Plains Air Force Base)	Lubbock, Texas	8XS8	84	32
Santa Fe Municipal Airport	Santa Fe, New Mexico	KSAF	286	74
Sidi Slimane Air Base	Sidi Slimane, Morocco	GMSL	77	29
Sondrestrom Air Base	Kangerlussuaq, Greenland	BGSF	44	28
Spangdahlem Air Base	Binsfeld, Germany	ETAD	20	10
Tyson Landing Zone	Yuma Proving Grounds, Arizona	--	15	15
Vicksburg Municipal Airport	Vicksburg, Mississippi	KVKS	108	66
Wake Island Airfield	Wake Island	PWAK	62	21
Waterways Experiment Station Asphalt Test Section	Vicksburg, Mississippi	--	96	25
Westover Air Force Base	Chicopee, Massachusetts	KCEF	74	23
Wheeler Air Force Base	Wahiawa, Hawaii	PHHI	61	17
Wilde-Benton Landing Zone Fort Bliss	Orogrande, New Mexico	--	11	11
Zaragoza Air Base	Zaragoza, Spain	LEZG	67	15
TOTAL			**4592**	**1533**

* Cases with numeric CBR values only.

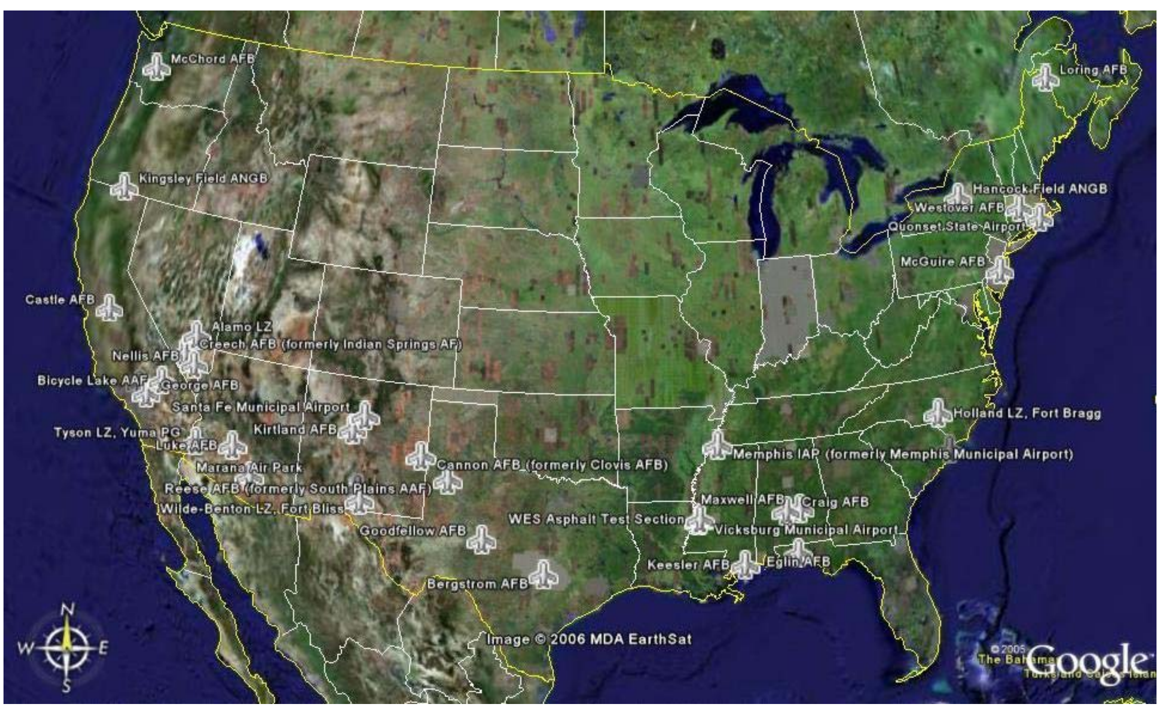

Figure 3.2: Geographic Distribution of Continental United States (CONUS) Test Sites.

Figure 3.3: Geographic Distribution of Pacific Area Test Sites.

Figure 3.4: Geographic Distribution of European Area Test Sites.

Table 3.3: Distribution of USCS Soil Types in CBR Database and Subset.

USCS Soil Classification	***Total Cases***	***CBR* Cases***
GW	71	42
GP	120	57
GM	214	115
GC	160	101
SW	101	29
SP	284	78
SM	807	245
SC	466	180
ML	304	77
CL	892	192
OL	0	0
MH	190	21
CH	205	89
OH	1	1
Pt	0	0
CL-ML	36	11
GW-GM	105	56
GW-GC	1	1
GP-GM	97	52
GP-GC	2	1
GC-GM	5	0
SW-SM	75	39
SW-SC	0	0
SP-SM	333	104
SP-SC	0	0
SC-SM	67	34
Missing	56	8
TOTAL	**4592**	**1533**

* Cases with numeric CBR values only.

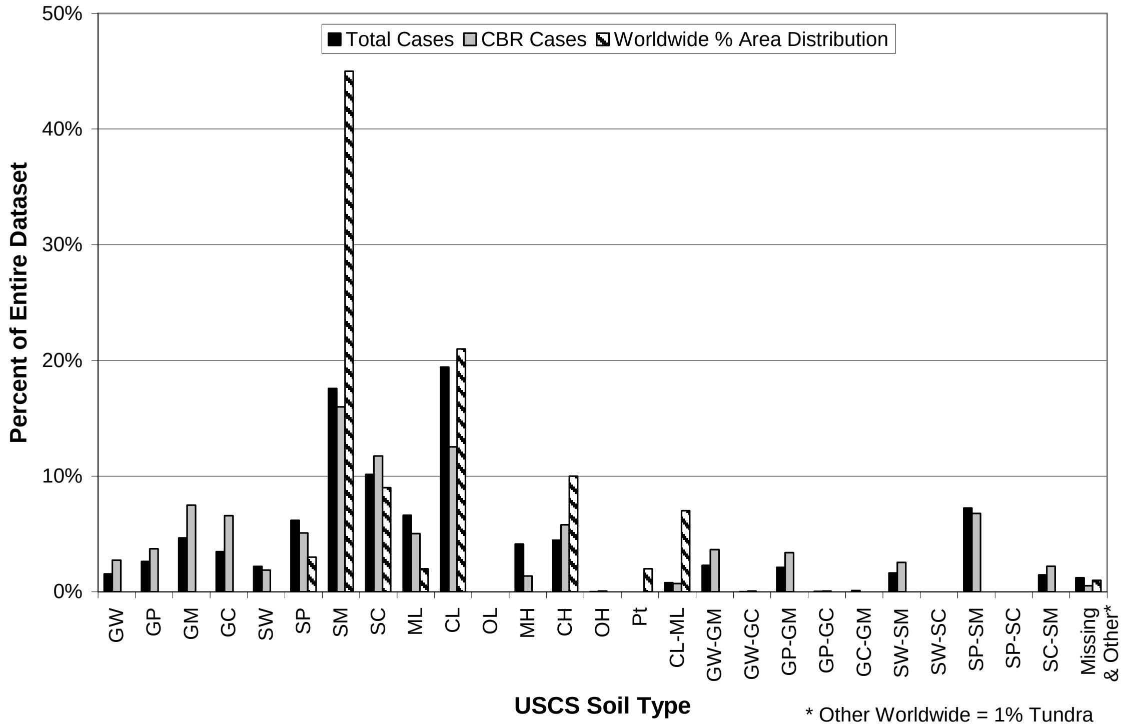

Figure 3.5: Distribution of Data by USCS Soil Type Compared to Worldwide Estimate [92].

3.4 Descriptive Statistical Summary

In order to provide a sense of the relative completeness of the data fields and their statistical distribution, summaries for both the full database and the numeric CBR subset are provided in Table 3.4 and Table 3.5. These contain only the numerical fields from the database and include only those that held significant amounts of unique data (i.e. none that were empty or contained unvarying parameters that were common to all or most entries).

The range and distribution for the features were compared to existing references in the literature to get a sense of how well the data represented what might be expected for naturally-deposited soils. One report that proved particularly valuable for this task was a compilation on the “Statistical Analysis and Variability of Pavement Materials” by

Freeman and Grogan [93]. This source collected numerous reports of statistical averages and distributions for most of the parameters listed in Table 3.4 and Table 3.5, including those for natural soil deposits. The plasticity characteristics (LL, PL, PI), specific gravity, optimum moisture content, maximum dry density, moisture content, dry density, percent passing the #4 & #200 sieves, and CBR statistics for the natural soils all compared favorably with the spread of the values in the database fields. Therefore, despite the reliance on collecting the data from airfield pavement testing, we felt reasonably confident that the database covered the range of conditions that one might expect to find in unimproved locations.

Table 3.4: Statistical Summary of Numeric Features in the Full Database.

Feature* (units)	Valid Cases	Quantiles					Mean	Standard Deviation	Coeff. of Var. (%)
		0%	25%	50%	75%	100%			
LL (%)**	1,999	14	23	30	44	85	34	14	41
PL (%)**	1,999	9	15	18	22	49	19	6	32
PI (%)**	1,998	1	6	13	21	53	15	10	69
SpGr	2,638	2.296	2.640	2.670	2.700	2.994	2.675	0.075	3
Depth* (in)**	4,592	0	11	20	34	90	23	16	70
OMC** (%)**	1,295	3.8	8.0	10.2	14.5	31.5	12.1	6.0	49
MDD** (lb/ft^3)**	1,343	89.0	112.5	124.5	131.5	151.0	122.1	12.7	10
MC (%)	4,020	0.5	5.8	10.8	17.1	85.3	12.8	8.8	69
DD (lb/ft^3)	1,686	64.5	104.3	116.2	128.7	168.7	116.4	16.2	14
3/4 M (%)	1,004	24	93	100	100	100	94	10	11
3/4 m (%)	1,004	24	71	90	99	100	83	17	21
#4 M (%)	1,817	12	68	96	100	100	83	22	26
#4 m (%)	1,817	10	53	86	99.5	100	76	26	34
#40 M (%)	1,004	4	33	60	91	100	61	30	49
#40 m (%)	1,004	4	20	35	76	99	46	32	68
#200 M (%)	1,838	0	14	32	54	100	38	29	76
#200 m (%)	1,834	0	6	24	50	100	32	30	93
0.005 M (%)	496	0	4	10	18	89	18	24	131
0.005 m (%)	496	0	0.25	2	8	75	10	18	178
0.001 M (%)	466	0	2	5	11	72	13	20	153
0.001 m (%)	466	0	0	0	5	57	7	14	207
CBR (%)	1,533	1	16	30	65	158	42.3	32.5	77

***** Key to abbreviations and acronyms used for features can be found in Appendix A.
****** Atterberg limits for cohesive soils only.
******* Depth below grade level.
******** Standard CE-55 compaction [17].

Table 3.5: Statistical Summary of Numeric Features in CBR-only Subset.

Feature* (units)	***Valid Cases***	***Quantiles***					***Mean***	***Standard Deviation***	***Coeff. of Var. (%)***
		0%	***25%***	***50%***	***75%***	***100%***			
LL (%)**	726	14	22	28	40	85	32	13	42
PL (%)**	726	11	14	18	22	47	19	6	31
PI (%)**	725	1	5	11	17	53	13	10	76
SpGr	1,088	2.296	2.640	2.670	2.700	2.994	2.669	0.079	3
Depth* (in)**	1,533	0	4	12	17	72	13	11	83
OMC** (%)**	698	3.8	7.8	10.0	13.9	31.5	11.2	5.2	46
MDD** (lb/ft^3)**	733	89.0	112.5	125.0	133.0	151.0	123.8	12.1	10
MC (%)	1,476	0.5	5.1	8.2	14.1	50.3	10.4	7.1	69
DD (lb/ft^3)	1,380	64.5	104.2	116.0	128.9	168.7	116.5	16.2	14
3/4 M (%)	526	24	89	98	100	100	92.5	11	12
3/4 m (%)	526	24	70	83.5	98	100	80	18	23
#4 M (%)	849	12	54	81	100	100	77	24	31
#4 m (%)	849	10	44.5	74	98	100	69	27	39
#40 M (%)	526	4	25	50	87	100	55	30.5	55
#40 m (%)	526	4	15	26	68	99	40.5	31	76
#200 M (%)	863	0	10	22	44	100	32	28	87
#200 m (%)	861	0	5	14	38	100	26	28	109
0.005 M (%)	269	0	4	9	18	89	15	20.5	135
0.005 m (%)	269	0	0	2	7	72	8	15	191
0.001 M (%)	257	0	2	5	9	72	10	17	164
0.001 m (%)	257	0	0	0	3	57	5	11.5	234
CBR (%)	1,533	1	16	30	65	158	42.3	32.5	77

***** Key to abbreviations and acronyms used for features can be found in Appendix A.
****** Atterberg limits for cohesive soils only.
******* Depth below grade level.
******** Standard CE-55 compaction [17].

3.5 Handling Difficult Data

3.5.1 Missing and Incomplete Data

The statistical summary presented in Table 3.4 clearly shows that a significant portion of entries for the 4,592 cases in the database were incomplete to some degree. In fact for the entire sixty-two fields in the database (Table 3.1), the "incompleteness factor" was a full sixty-five percent of all entries. However, as noted previously, the structure was set up more for flexibility than the intention of filling all records completely. A total of thirteen fields were completely unused and an additional eight contained data for less than five percent of all cases. Another twelve fields contained essentially descriptive or reference information that were generally not useful for prediction method analysis. Information for the fields containing landform, lithology of the soil parent material, method of soil deposition, depth to water table, and mineralogy of the clay fraction were gleaned from the text of the pavement evaluation reports or inferred by their geographic location. As such, the subjective nature of this data resulted in a low degree of confidence so it was not used in the data mining process (in addition to the difficulties of dealing with their categorical nature as discussed in section 3.5.2)[**]. Of the remaining twenty-six fields in the database that were not mostly empty, reference information, or subjectively assigned, the "incompleteness factor" improved slightly to an overall fifty-nine percent missing data. The depth below grade, USCS soil type, moisture content, and pavement layer were the most complete features ranging from one hundred

[**] Another problem exists with using these features for modeling. Although geomorphological factors influence the formation of different soil types, direct linkages are difficult to establish [94].

to eighty-four percent of cases. The fine particle sizes (0.005 and 0.001 mm) were the least complete features, only containing data in approximately ten percent of cases[††].

Clearly, even cases with missing data contain some degree of useful information that can be helpful in establishing relationships between inputs and outputs when building models. With a significant number of incomplete cases in the database, we hoped to find ways to incorporate them into the data mining process. However, it has been acknowledged that "the problem of missing data is very complex…[and] much research remains to be done" [13]. There are several common methods for dealing with missing data [95]. The simplest is case deletion, where only the cases with complete data are used. Imputation is another approach, which tries to estimate missing data by relating it to other variables that are available for that case. Other techniques, such as maximum likelihood estimation and Bayesian methods, are much more complex and require some assumptions regarding the distribution of the data that can be hard to meet. One novel approach for neural networks is based on an alternate idea. Instead of trying to infer a replacement for the incomplete data, the connections and neurons that encounter the missing features are temporarily switched off during the training process [96]. By associating each case in the database with a corresponding "validity vector" of binary numbers that identify legitimate versus missing data, the network can adapt its architecture each time data is absent. The developers actually found that networks trained this way with incomplete data were better at making predictions for cases with missing data than networks trained with complete information for all records.

[††] However, this would be expected as these are typically measured only for fine-grained soils, such as silts and clays, where there is a significant portion of fine material present, and requires a hydrometer analysis in addition to a sieve-size analysis.

Due to the substantial complexity of dealing with missing and incomplete data and the introduction of biases that can result with many of the common methods, a decision to deal only with complete cases was made. However, in many instances this was not as limiting as it seems because it did not preclude the use of pairwise analysis (using all cases where both features are present) and varying the size of data subsets when fewer features were in use. Initially we had hoped that some of the artificial intelligence methods would implicitly handle missing data. Unfortunately, these techniques are not a panacea in dealing with the complexities of missing information.

3.5.2 Categorical Data

Most analytical methods are primarily suited to handle numerical data, so dealing with categorical features can pose special challenges. Categorical data represent discrete values and there are three types: binary, ordinal, and nominal [97]. Binary data takes on one of two possible values (e.g., yes-no, true-false, etc.), ordinal data represents categories that have an ordered or hierarchical relationship (e.g. A-B-C, or Monday-Tuesday-Wednesday), while nominal data denotes classes that have no implicit order among them. The difficulty arises in deciding how to map these nonnumeric data into a numeric form, while preserving the underlying meaning of the categorical features. There are several common rules for this, which involve different approaches depending on the type of categorical data [98].

For features where the data is binary, it can be mapped to some combination of the values of 0 and 1, or 1 and -1. One goal, while performing this transformation, is to keep the average value near zero. So, when one binary category dominates the dataset or represents a rare event, the use of 0 and 1 is common. The dominant and “regular” case

should be assigned to 0, while the rare event should be coded as 1. When both of the categories are equally probable (roughly), the data should be mapped to the values 1 and -1 to preserve the zero average balance.

In cases where categorical data have an ordinal relationship, there are two mapping strategies. The first strategy makes sense when there is some relationship among the classes where each class represents a particular value. In this case, one variable that takes on the corresponding value for each case may be used. An example would be a typical academic grade scale, where an A is assigned a score of 100, B a score of 90, C a score of 80, etc. The second strategy should be used when there is an order to the categories, but no inherent value is associated with them. In this case, n-1 binary inputs are used to represent the n distinct categories. For the lowest valued class, all of the n-1 inputs are set to a value of -1. For the next higher valued class (i.e., second lowest), the first of the n-1 inputs is given a value of 1, while all the others remain at -1. For each subsequent higher class, another of the n-1 inputs is given a value of 1, and this pattern continues until the highest class is reached. In the case of the highest class, all of the n-1 input values are set to 1. Despite requiring the use of considerably more variables, “many nonlinear models respond well to this encoding” [98].

For nominal data, where the categories have no inherent order, the following strategy is commonly employed. To represent each of the n classes, one binary variable is used for each class. For each of the n categories, one of the variables is set to 1 while the rest are kept at 0. Because this requires the use of many variables in a model, one may be tempted to simplify this approach. However, there is a strong warning to “*Never* encode unordered classes by using different values of a single variable” [98].

A limited number of analytical techniques[‡‡] are able to deal with symbolic features directly, without the transformation to a numerical representation. The nearest-neighbor method can use the hierarchical relationships among ordinal classes in determining similarity among cases [99]. Decision tree methods (a.k.a. recursive partitioning) can base the splitting of the data into separate groups directly on symbolic feature values [99]. Also, decision trees are naturally suited to dealing with relatively large numbers of binary or categorical attributes, provided that the number of categories is small [100].

The need to limit the number of input features to a reasonable level poses a particular problem for ordinal and nominal data. Given the common strategies for representing these symbolic features with numerical data, they can result in an explosion of model inputs. With a finite number of cases in the database, increasing the dimensionality of the problem domain quickly leads to an inadequate and unrepresentative coverage of the entire space. This phenomena is known as the “curse of dimensionality,” which describes the problem that “when many attributes are involved, the amount of data that you have *relative to the entire space of possibilities* is very small” [99]. In fact, as the number of attributes increase, the volume of the decision space grows exponentially [101]. In addition, more attributes place increased demands on the computational power needed for model-building.

Much of the symbolic data collected for the OLS database had large numbers of categories (USCS soil type for example) or were assigned their values rather subjectively (as discussed in the previous section). Because of these reasons, and the other

[‡‡] See sections 4.1 and 4.2 for a detailed discussion of specific prediction methods.

representational difficulties described above, the decision was made to concentrate modeling efforts on numerical features only.

4 Approach

4.1 Choosing Prediction Methods to Evaluate

In the absence of a clear understanding of the complete problem domain and the inherent characteristics of the dataset available, the choice of an appropriate system model a priori can be difficult. For most "real-world" problems (i.e., not artificially generated) the most widely recommended approach to choosing a prediction method is trial and error [102]. From an empirical comparison of the results of candidate methods applied to the problem at hand, the best performer can be chosen. Depending on the relative importance of a particular performance metric for an application, this decision may be made on the basis of error rates, speed of learning, or the degree of understanding they provide about the decision-making process.

Some attempts have been made to determine which techniques work best on certain classes of problems. The StatLog project [100] conducted a comparative study of about twenty different statistical, machine learning, and neural network classification algorithms on approximately twenty large-scale, industrially-important, real-world datasets. The aim of this comprehensive evaluation was to characterize the suitability of methods for different applications based on various descriptive characteristics (metadata) of their datasets. Unfortunately, it appears that the resulting rule-based expert system that scores the methods for new applications based on their dataset metadata is no longer available for use, even though the datasets it is based upon are [103]. However, some generalizations regarding which methods work best for certain application areas are provided by the StatLog study. For problems where the task involves modeling a human decision-making process (e.g. credit risk assessment), decision trees seem to perform

best. For problems involving image data, nearest neighbor methods generally excel, while a further breakdown into object recognition versus image segmentation tasks reveals that neural network and standard statistical procedures do well for the former while decision trees are better for the latter. For problems involving the assessment of cost or risk, the study suggests that some types of neural networks may perform poorly, but they might be best for problems involving prediction or partitioning.

Another effort that allows comparisons of learning methods on a common set of databases is the DELVE project (Data for Evaluating Learning in Valid Experiments). "DELVE is a standardised environment designed to evaluate the performance of methods that learn relationships based primarily on empirical data. DELVE makes it possible for users to compare their learning methods with many other methods on a wide variety of datasets. The DELVE learning methods and evaluation procedures are well documented, such that meaningful comparisons can be made" [104]. There are presently twenty-five learning methods and eighteen datasets available within the DELVE environment, however it appears that no systematic approach to testing all applicable methods on each dataset has been undertaken. Even though no overall conclusions regarding method selection can be currently drawn from this initiative, it serves as a useful benchmarking resource for checking that customized implementations of a particular method perform comparably to established results.

An alternative concept for selecting the optimal analysis method for a specific problem is the Intelligence Density Framework (IDF) technique proposed by Dhar and Stein [99]. The IDF system embodies the classic systems approach to problem solving, where a set of alternative solutions is evaluated against the problem's specifications. In

particular, they recommend comparing the relative strengths and weaknesses of each individual analysis technique under consideration with the objectives and constraints of the problem domain. Methods with “Intelligence Density Profiles” (IDP) that most closely match the IDP of the problem are considered to be the best candidates for providing good solutions. The specifications used in making this comparison involve fifteen metrics that describe the problem domain. These “dimensions” are listed in the left-hand column in Table 4.1, broken down into four major groups that consider application demands, method implementation, dataset characteristics, and human interaction separately.

Looking at this decision matrix, the profiles for the methods can be compared to an estimation of requirements for the Opportune Landing Site soil strength prediction task. Fuzzy logic is clearly the one technique that stands out as the best among the group. However, fuzzy set theory is really just an approach to representing imprecise numbers and not a prediction method in and of itself. It is useful in capturing the “vagueness” of natural language and converting this into a number, but fuzzy logic must be combined with another method to form a complete modeling system [105]. For example, fuzzy numbers may be incorporated into statistical regression [106], rule-based systems [107], or neural networks [108]. As a “first cut,” it makes the most sense to test prediction methods alone and choose the best performing one before proceeding to hybridized approach. So, even though it ranks as the best match for the soil strength problem, the use of fuzzy logic was put aside during this study until the best stand-alone prediction method could be selected.

Table 4.1: Intelligence Density Profiles for Several Candidate Methods and the Current Problem (based on the framework of [99], format after [109]).

Dimensions sorted by GROUP	*Genetic Algorithms*	*Neural Networks*	*Rule-Based System*	*Fuzzy Logic*	*Case-Based Reasoning*	*Decision Trees*	***OLS Soil Strength***
APPLICATION DEMANDS							
Accuracy	3	5	3	5	4	4	**5**
Explainability	2	1	5	3	3	4	**3**
Response Speed	4	5	2	5	4	5	**3***
METHOD IMPLEMENTATION							
Scalability	3	3	3	3	5	4	**4**
Compactness	3	5	1	5	2	3	**4**
Flexibility	5	5	3	5	5	5	**5**
Embeddability	5	5	1	3	3	4	**4**
Ease of Use	3	3	3	3	3	3	**3**
DATASET CHARACTERISTICS							
Learning Curve	5	5	2	3	2	2	**2**
Tolerance for Complexity	5	5	1	5	3	3	**5**
Tolerance for Noise in Data	3	4	--	--	3	3	**5**
Tolerance for Sparse Data	3	1	--	--	3	1	**5**
HUMAN INTERACTION							
Independence from Experts	5	5	1	3	4	3	**4**
Speed of Development	4	3	4	3	4	3	**3**
Computing Resources	2	2	3	1	2	3	**2**

Numerical scale represents strengths and weaknesses from best (5) to worst (1).

-- Signifies that this dimension is not applicable to the given method.

***** When used as a decision support tool for planners. If deployed as an airborne system that allows pilots to pick OLS in real time, the response time would need to be almost instantaneous.

Comparing the IDP of rule-based systems to the requirements for the OLS task reveals that this method stands apart as the worst. This makes sense because rule-based systems are based on expert knowledge and work best for problems where these rules can

be articulated clearly [102]. Obviously, this is not the case for the soil strength prediction problem before us here.

The four remaining prediction methods listed in Table 4.1—genetic algorithms, neural networks, case-based reasoning, and decision trees[§§]—all cluster at the middle of the pack in comparison to the OLS task. Genetic algorithms, which are also known as evolutionary computing, focus on a "survival of the fittest" strategy when it comes to adaptation and learning. With a powerful capability to optimize, "genetic algorithms are very good at [feature selection], having a capability to search through large numbers of combinations where there may be interdependencies between variables" [110]. But, genetic algorithms are not a stand-alone prediction method, so like fuzzy logic the approach must be paired with another technique. This leaves neural networks, case-based reasoning and decision trees as the best candidates that can be evaluated individually.

Several sources [100,102] recommend that any basic "toolbox" of good methods includes all three of these plus linear regression. They suggest trying one of each type from the four classes based on the following reasoning. Linear regression serves as a "classic" standard algorithm that is widely-available and applied. As a more "modern" statistical approach, nearest neighbor methods often work well and allow a nonparametric analysis, which does not rely upon underlying assumptions regarding data distribution. Decision trees are computationally cheap, allow easier interpretation of their decision methodology, and appear to complement nearest neighbor algorithms – each tends to

[§§] Case-based reasoning and neural networks are described thoroughly in sections 4.2.2 and 4.2.3. The nature of decision trees methods can be summarized as follows: They are based on a recursive, binary partitioning of the available database into smaller and smaller groups. The splitting criteria are established by determining which one value of a single input variable produces the greatest degree of statistical difference among the cases in the two groups that result.

perform well when the other does not. Finally, neural networks provide a powerful technique with the potential advantage of abstraction in larger dimensions.

During the testing phase of this work, some initial trials with decision trees did not yield very promising results. Because of the limited time available, the choice was made to focus testing efforts on the remaining three of the four methods. The linear statistical regression was not expected to perform well either, but it is straightforward and provides a good baseline for comparison. The nearest neighbor and neural network approaches received the most attention in the evaluation phase of this study.

4.2 Candidate Prediction Methods

In general, there are two different but related types of prediction tasks [111]. The distinction lies in what type of dependent variable is predicted with the independent variables: a discrete categorical class or a continuous numerical value. In the case where the prediction result is a category, the task is called classification. For problems where the prediction is a number, the task is called regression. Although the same tools can be used for both tasks, there are slight differences in how they are applied. For example, when error is measured for a classification model the result is either correct or incorrect, and a misclassification rate is used. However, in regression the magnitude of the error is important and a least squares approach is used [102].

For the OLS project, the goal is to predict the CBR value for soil strength. Predictions of soil strength classes might be carried out, but ultimately the numerical value for CBR is needed to assess the aircraft types, payloads, and coverages that the soil can support. Therefore, this study focused on prediction methods as they apply to a regression task.

4.2.1 Statistical Regression

The simplest and most traditional approach to making predictions is to employ classic statistical regression techniques. With these methods, we choose the form of a function a priori and then optimize its parameters to achieve the best fit to the data. The most basic flavor when more than one input feature is involved is called multivariate linear regression. This relationship takes the following form:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n \tag{4.1}$$

where:

$x_1, x_2, \ldots, x_n$	are the input features (independent variables, regressors)
y	is the target (dependent variable, response)
$\beta_0, \beta_1, \beta_2, \cdots, \beta_n$	are the regression coefficients

Combinations of input features can be used or the regressors can be nonlinear themselves, but the relationship is considered to be linear when it is linear with respect to the beta parameters [112]. Other nonlinear relationships can be represented by first transforming the data so it exhibits a linear trend. Also, normalization of the input data so the parameters are scaled equally can allow an assessment of the linear sensitivity of each input on the output.

In the case when the regressors are linear and each is paired with a single beta coefficient, the process can be thought of as fitting an n-dimensional hyperplane to the data. The regression coefficients are determined by the method of least squares. For the case of a simple linear regression for one variable, the method is illustrated in Figure 4.1, where the regression is a line. The two beta parameters, which represent the slope and

the y intercept for the line, are chosen to minimize the sum of squared distances from the regression line to each of the data points.

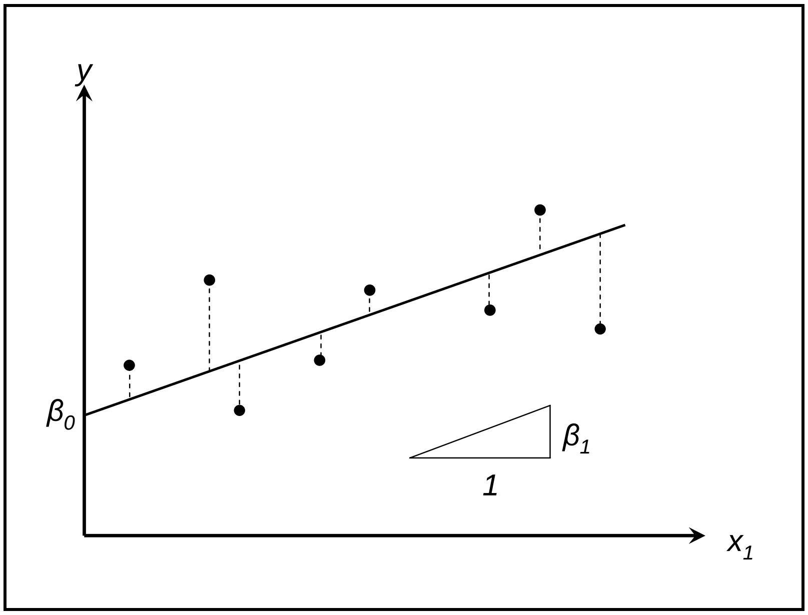

Figure 4.1: Least Squares Fit of a Regression Line to Observed Data Points for a Single Variable.

More sophisticated approaches to statistical regression are available. These include using more than one line or curve in different areas of the dataset to generate more complex surfaces, a process known as piecewise regression. The choice of the regressors to use in a model can also be varied to find the best combination for good performance. One can start with a single variable and find the best new ones to include, or begin with all variables and take away the least important. This approaches are known respectively as forward and backward stepwise regression.

A drawback to statistical regression for complex problems is deciding how to choose the form of the function to fit the data with. "If the relationship between x and y is non-linear, regression analysis can only be successfully applied if prior knowledge of the nature of the non-linearity exists….In the real world, it is likely to encounter

problems that are complex and highly non-linear. In such situations, traditional regression analysis is not adequate" [113]. One benefit of linear regression is that when our goal is to find a solution that minimizes errors on new cases, simple methods tend to do well because there is no danger of overfitting the data used to build the model [102].

4.2.2 Knowledge-Based System Approaches

Knowledge-based systems involve methods that use a body of prior experience in order to make new predictions. "Knowledge-based systems (KBS) are suited to developing applications which make use of heuristics, or empirical knowledge, rather than solving a set of mathematical equations" [12]. There are two major categories of knowledge-based systems, each with a different focus on what kind of information is captured and then used to make new decisions. Rule-based systems concentrate on the logic behind the decision-making process itself. Case-based systems work with a collection of examples and make decisions based on those that are most analogous to the current problem.

Because they are based on the reasoning of skilled practitioners in a specific problem domain and try to mimic their performance, rule-based systems are commonly known as "expert systems." Rule-based approaches use a set of simple logical rules, which can be chained together to achieve complex reasoning [99]. These systems are built by interviewing human experts to determine explicit rules, which are then encoded into a computer model. This approach is particularly suited to problems where the amount of available data is limited, but the decision-making process can be articulated clearly by someone familiar with the application domain [102].

“Case-based reasoning is an alternative to rule-based reasoning. Here, the new problem to be solved is matched with existing cases in the knowledge base….Similar cases can therefore be recalled, and decisions for the new problem can be based on what has been done before” [12]. In effect, the collection of past cases itself becomes the model and thus is easily updated when new information becomes available. And in this manner, the system can “learn” over time. As new examples are added to the case base, better answers result. Solutions are created by “synthesizing the similar cases and adjusting the final answer for differences between the current situation and the ones described in the cases” [99]. Because it is based solely on examples, case-based reasoning is popular in areas where there are no clear-cut theories to explain the phenomena [99].

One approach to selecting the most similar cases to make a prediction from is known as “*k*-nearest neighbor” (*k*-NN). This technique involves measuring the distance between the input vector that we want to make a prediction for and all of the others in the existing knowledge base to find the *k* closest. For *n* continuously-valued numerical features, this can be thought of as searching an *n*-dimensional space to find the closest points to the one we want to predict. Different distance metrics can be used to determine how close cases are to the new example. The most common technique, illustrated in Figure 4.2, is the Euclidean distance between points, but others are available such as the city-block distance between points or the cosine of the angle between them [110]. The cosine distance metric can be problematic with points that are collinear with respect to the origin.

Though it is more frequently used for classification applications, *k*-NN can also be employed for regression problems. The key is what we do with the target values once the *k*-nearest neighbors are found. In the case of classification, the most common class for the majority of cases is usually used as the prediction. (For this reason, *k* is usually odd in classification tasks to avoid ties.) With regression, however we must decide how to treat the numerical target values of the *k* neighbors [110]. The simplest and most widely-used approach is to take the average of the *k* values as the prediction. However, this does not take into account the relative distance of the neighbors from the point we want to predict. Because the entire method is based on the reasoning that the most similar cases should make the best predictors, it makes sense to apply a relative weighting to the neighbors based on their distance to the point of interest. Some examples of distance weighting schemes include using the inverse square of the distance (equation 4.2) and the negative exponential distance (equation 4.3).

$$W_i = \frac{\frac{1}{d_i^2}}{\sum_1^k \frac{1}{d_i^2}} \tag{4.2}$$

$$W_i = \frac{e^{-d_i}}{\sum_1^k e^{-d_i}} \tag{4.3}$$

where:

W_i	=	Weighting factor for the i^{th} of the *k* neighbors
d_i	=	Distance from the i^{th} of the *k* neighbors to the prediction point

Note that the inverse squared distance method can be problematic if a “neighbor” overlaps the point that we wish to predict because of division by zero. A third task that the *k*-nearest neighbor method can perform is to serve as a decision-support tool to provide the most similar cases to an expert for further analysis [99].

The choice of the *k* parameter is important and influences how the method performs. In effect, *k* can be thought of as a smoothing parameter [110]. If k is small, then the variability of the predictions will be high, but if k is large then the predictions will be based on points that are far away from our point of interest. A value for *k* is usually determined empirically, by stepping through a series of *k* values and determining which provides the best performance.

The *k*-nearest neighbor method has the advantage that it is completely nonparametric, meaning that no assumptions are made with regards to the statistical distribution of the data [102]. As such, it can produce any arbitrarily complex surface in the problem domain. *k*-NN is also well-suited to handling discontinuities in the relationships between variables [99]. The nearest neighbor method is also computationally inexpensive for the “learning” phase of implementation, since this merely involves building a case base.

However, the disadvantage in computational power comes when predictions are made, since similarity distances must be computed for every example in the case base [102]. In addition, since the entire knowledge base needs to be available to use the method, it is not particularly compact or portable when incorporating into larger applications [99]. Other drawbacks to the *k*-NN method include its sensitivity to the relative scaling of the features, so close attention to weighting them properly or

normalizing is important [99,102]. The *k*-nearest neighbor approach is sensitive to the number of features used as inputs. As the dimension of the problem space increases, computing similarity becomes difficult because the “neighborhood” becomes very large and neighbors are spaced far apart [110]. Also, *k*-NN does not work well when there are interactions among the variables, especially subtle ones among continuous variables [99].

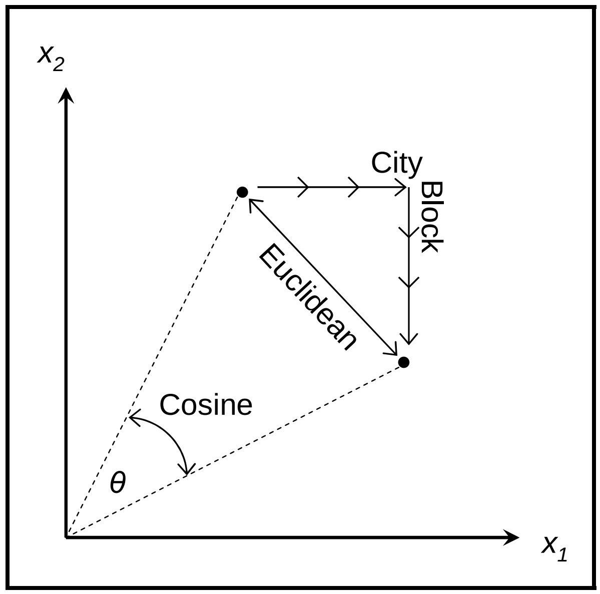

Figure 4.2: Distance Measurements for Nearest Neighbor Methods.

4.2.3 Artificial Neural Networks (ANNs)

Artificial neural networks are powerful learning methods that are structured and operate in ways that are analogous to biological nervous systems [101]. They consist of a collection of simple processing units, called neurons, which can be joined together to form network structures. A single neuron is illustrated in Figure 4.3. It accepts a series of input values (x_m), which are multiplied by weighting factors (w_m) before being summed together in the neuron itself. One of these weights, called a bias (b), is special in that it is always receives a constant input, which serves to offset the output of the

neuron. The weighted sum of the input and weight products (y) then becomes the input to an activation function (φ). Common forms of activation functions include hard-limit (i.e., step functions), sigmoid shape, radial basis (smooth bell-shaped) and linear [114]. Typically these functions are bounded from 0 to 1 or -1 to 1, so neurons work best on data scaled to this range to prevent saturation of their output [13]. The result from the activation function is the final output of the neuron.

A number of neurons can be joined together in layers to form a network, as illustrated in Figure 4.4. Inputs are connected to the first layer of neurons by a series of weights. The outputs from the first layer neurons serve as the inputs to subsequent layers by way of another series of connection weights. The first layer of neurons and any other layers that precede the final layer are known as hidden layers. (In this example there is only one hidden layer, but more can be used.) The final layer of neurons is known as the output layer and the result from these serves as the output of the entire network. The number of input nodes and output neurons are determined by the dimensions of the input and output vectors of the application. In this example, the network is fully-connected in the sense that each input node and neuron is connected by a weight to all of the following neurons. Networks with fewer connections are possible. The network is also a feed-forward type, where information flows from the inputs to the outputs only, with no connections going from one neuron to previous layers—however, this too is possible for some neural networks.

Neural networks "learn" by a process of adjusting their weights and biases to optimize the best mapping of inputs to outputs. This process can be supervised, where the outputs are compared to known, correct output values (targets), or unsupervised,

where there are no targets. Unsupervised learning can be helpful in finding structure in the data, such as clustering or patterns, however supervised learning is most appropriate to the regression task of fitting a function to known target values [101]. To start, the weights and biases are assigned random values, and then the inputs are presented to the neural network. For each input case, an output is computed, compared to the target value, and an error value recorded. In batch training all of the input cases are presented before any weights are changed, while sequential learning involves correcting the weights and biases after each presentation of an input vector. For batch training, errors for all cases are combined to compute an overall measure of the error for the entire training set. Mean square error is a popular error function.

In order to minimize the error by adjusting the weights and biases, the back-propagation technique is used. In this concept, the gradient of the error function is computed with respect to the weights and biases. (For this reason, activation functions for neurons must be differentiable.) Then the weights and biases are updated in a direction that steps towards a lower value on the error function surface. Many different learning methods are used for deciding how to update the weights after each new batch presentation of the inputs, with an eye towards faster convergence to the minimum error. The direction to move and the distance must be determined. Methods for picking a direction include using the negative direction of the steepest gradient (simplest), using a conjugate (orthogonal) gradient relative to the previous step, and Newton-type methods where the gradient is recalculated at each step. Methods for changing step size include steadily decreasing it over time, using “momentum” terms that allow past steps to contribute by helping avoid local minima, or allowing adaptive sizing based on whether

the error rate is improving or getting worse at each step [114]. The process of presenting the inputs to the network, recomputing the error, and updating the weights and biases to improve performance continues until the desired model precision is reached or the error gradient flattens out and does not change (where hopefully a global minimum error has been reached).

Additional stopping criteria are used to prevent the network from relating inputs to outputs so well that it begins to map the noise in the training data. This phenomena is known as “overtraining,” which results in an overspecialized network that performs superbly on the training cases but will not generalize well to new cases it has not seen before. Two approaches to avoid overtraining include early-stopping and regularization [114]. For early-stopping, the data used to build the model is split into two subsets: the “training set” used to train the network and a “validation set” used to monitor its performance on unseen cases[***]. While the weights and biases are iteratively adjusted to drive the error on the training set lower and lower, the error on the validation set is monitored. Initially, when the network is poorly trained, the error rate on the validation set is high. As training continues, the validation set error rate declines but then starts to increase once overtraining begins to occur. Training is stopped when the error rate on the validation set reaches a minimum value. Regularization is a different approach to judging the optimal amount of training. In this technique, a “penalty” term representing the complexity of the neural net is added to the network error function. As training proceeds, the network weights and biases are updated to minimize this regularized error

[***] A third subset of data independent of the model building process, termed the “test set”, is needed to accurately estimate the generalization error for new cases. This set is used after the model is completely designed using the training and validation sets. Evaluation with an independent test set is applicable to all prediction methods, not just artificial neural networks.

function, thus balancing low error and high complexity. With regularization, a validation set is not needed so all of the model-building data can be used for training.

The representational power of a neural network is due to its hidden layer neurons. Nonlinear activation functions in these layers permit networks to map nonlinear relationships between inputs and outputs [13]. It has been proven that one hidden layer of neurons with continuous activation functions can approximate any continuous function and two hidden layers can approximate any function, both to any desired degree of accuracy depending on the number of neurons in each layer [115,116]. The number of neurons in these hidden layer(s) must be selected carefully. Too few hidden units can result in a network that does not have the flexibility to map inputs to outputs if the relationship is complex. Too many hidden units can result in a network that is so powerful at relating inputs to outputs that it begins to map the noise in the data it is trained with. This is another way in which networks can become overspecialized and fail to generalize well to new data. Similar to using a validation set to avoid overtraining, another subset of the data called the "test set" is used to judge what size network is best. The test set contains data that is put aside and not used in the model-building and training process, so it is a truly independent dataset that can be used to reliably estimate generalization error for unseen cases. The number of neurons in hidden layers and the number of hidden layers are determined empirically by using the test set error to optimize them. Some rules of thumb regarding the network size relative to the number of training cases exist. Guidelines based on the number of weights and biases in a network are more reliable than those involving solely the number of neurons [13]. A recommended rule

involves trying networks with a total number of weights and biases between one-half to one sixty-fourth of the number of training cases [117].

Artificial neural networks are particularly suited to problems where there is an abundance of data available and explicit models are difficult to formulate [102]. They excel in applications that demand nonlinear and multidimensional mapping of inputs to outputs [101]. Prior knowledge regarding the nature of these complex relationships is not required [113]. ANNs are tolerant of noisy data and generalize well within a problem domain where the data may be poorly-distributed or spotty [99]. Once trained, neural networks can compute predictions quickly with a high degree of accuracy and are easily incorporated through simple algebraic equations into larger applications [99].

The main criticism of artificial neural networks is that the reasoning behind the input-output relationship mapping is opaque, leaving us largely unable to follow and interpret how decisions are made. Some methods exist to analyze trained neural networks but they are generally involved, not robust, and can prove difficult to interpret even the relative significance of input features [99,118]. Combining the learning power of neural networks with other prediction methods can result in hybrids that can alleviate some of these problems [108,119]. Another issue with ANNs is the design methodology tends to be largely empirical, so finding the best combination of parameters (network type, connection architecture, activation function, learning method, number of layers & neurons, etc.) for a specific problem can be very time consuming and require significant expertise [99]. Finally, finding an optimal solution that reaches the global minimum error for the problem is not guaranteed with a neural network, though in many cases they do outperform more conventional methods [113].

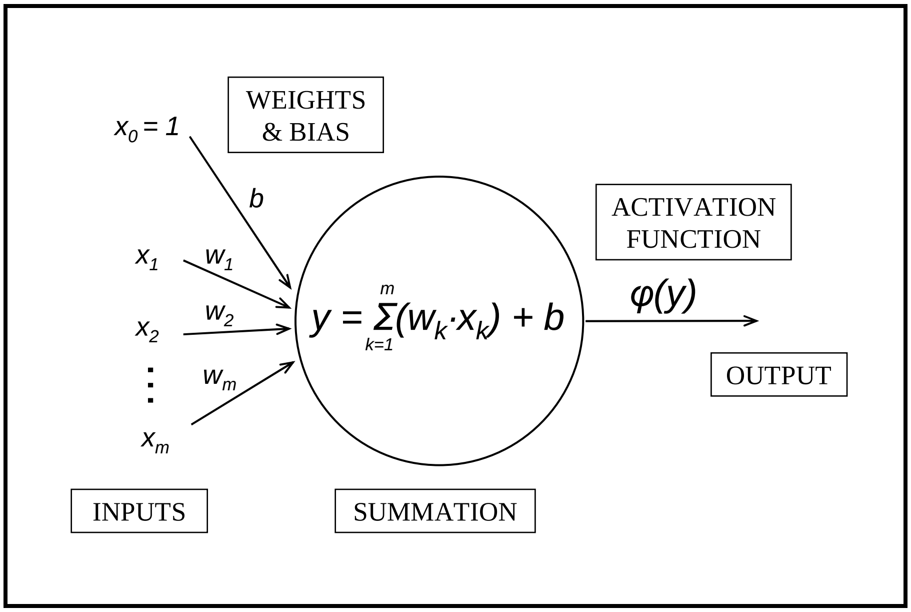

Figure 4.3: Diagram of a Single Artificial Neuron.

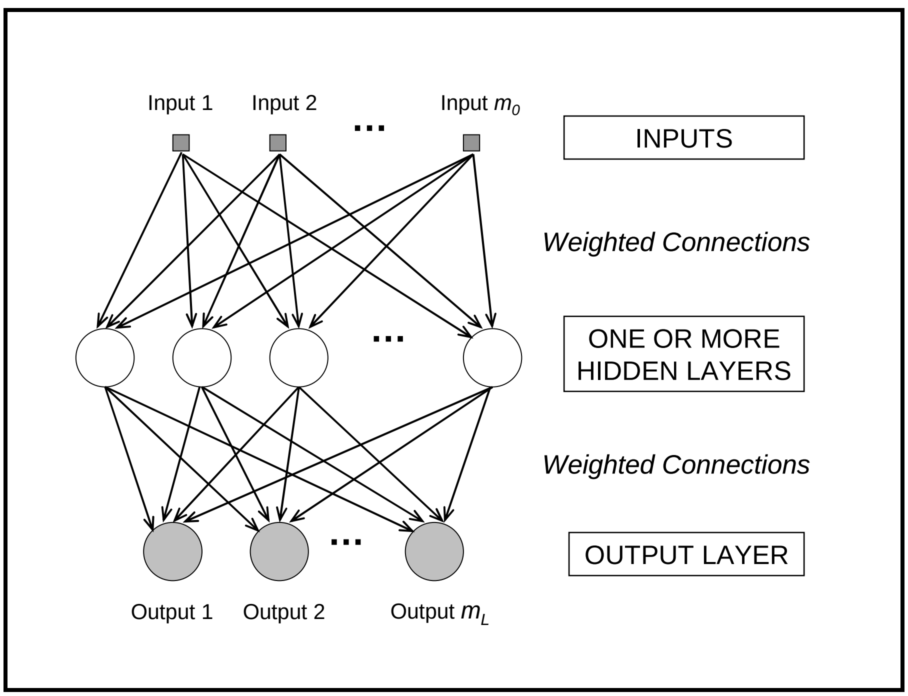

Figure 4.4: Structure of a Fully-Connected Feed-Forward Artificial Neural Network.

4.3 Feature Selection

From surveying the literature (see section 2.2) it is clear that many factors influence soil strength. In fact, "the CBR value has been correlated with some fundamental properties of soils, such as plasticity indices, grain-size distribution, bearing capacity, modulus of subgrade reaction, modulus of resilience, shear strength, density, and molding moisture content" [120]. Among the most frequently cited are:

- Moisture Content [6,8,18,33,39,40,42]
- Density [8,18,33,38–40]
- Plasticity Characteristics [8,25,32,38–40,42]
- Amount of Fine Particles [8,18,25,32,43]
- Gradation [8,25,32]
- Optimum Moisture Content [6,39,41]
- Maximum Dry Density [41]
- Particle Shape [8,43]

From a general overview of these sources, the most important factors appear to be moisture content, density, and plasticity. Using this knowledge of the problem domain as a firm practical basis, we can proceed to more quantitative methods to help select the best features from the OLS CBR Database for modeling analysis.

From the numerical data fields in the full database, a correlation matrix (shown in the upper half of Table 4.2) was used to assist in the elimination of redundant features and to determine which were most highly correlated with the CBR index. The Pearson product-moment correlations (*R*) among all the features and the target CBR value were calculated using the JMP statistical software package [97]. Each entry in the matrix

represents the correlation coefficient between two features. The term indicates the strength and direction of a linear relationship between the two features, with 0 indicating no relationship, 1 and -1 revealing a perfect linear correlation (increasing or decreasing, respectively).

Due to the degree of missing entries in the database, a pairwise approach to calculating the correlation coefficients was adopted. In this manner, every case in which data exists for both features in a single pairing is used to determine the value. So from pair to pair the number of supporting data used to calculate the correlation coefficient varied, in some instances quite widely. The number of data for each pair of features is given in the lower half of Table 4.2.

Almost all correlations were significant at a level of 0.05, meaning there is no greater than a 5% probability that the value of the correlation between those features is due to a random chance. The only exceptions to this were correlations between field dry density (DD) & specific gravity (SpGr), the maximum value for percent passing the ¾ inch sieve (3/4 M) & specific gravity, and the maximum value for percent passing the ¾ inch sieve & the plasticity index (PI).

In order to confirm that the correlation coefficients were reflective of the trends in the data, a matrix of bivariate scatterplots for all features and the CBR target was produced. Each one of these plots reflects the distribution of the data between two of the features (or one feature and the target CBR). The scatterplots indicated that the correlation coefficients were due to real relationships in the data, and not a spurious product of outliers.

Looking closely at each of the scatterplots for the input features versus CBR also validated the initial assumptions that CBR is not a linear function of any of the numerical features, but some general trends are apparent. CBR shows an inverse relationship with liquid limit, plastic limit, plasticity index, depth, optimum moisture content, gravimetric moisture content, and percent passing the #200 sieve size (Figure 4.5). CBR has an exponential relationship with both maximum dry density and field dry density (Figure 4.6). These make sense since we expect soils with higher plasticity and more fines to have lower strengths when wet, and higher density soils to correspond with higher strengths. Attempting to fit curves accurately to any of these relationships did not seem feasible given the spread in the data. Trying to fit an upper bound on the data seems possible for some features, but there is really little utility to this. In order to provide conservative predictions a lower bound is needed, but the inversely-related features appear to have none and the exponential data would produce one that is so low as to be useless. It was hoped that the nonparametric and nonlinear modeling power of the machine learning methods could be exploited to develop the best multivariate fit of the features to the CBR target.

Using the correlation matrix and some commonsense intuition regarding the physical meaning of the features, the following observations can be made:

- CBR does not show a strong linear relationship with any of the features. The highest correlations are with the coarser grain size particles with values of 0.5 to almost 0.6. This makes sense, since we would expect gravels to exhibit high strengths and provide the least complicated way for assessing whether a material will be strong or weak.

- Liquid limit, plastic limit, and plasticity index are highly correlated. This is expected as plasticity index is defined as the plastic limit minus the liquid limit. Therefore, we should eliminate one of these features from the model input. In this case, liquid limit appears to be the feature most correlated with the remaining two, so this feature could be removed[†††].
- Liquid limit, plastic limit, and plasticity index are highly correlated with the percent of fine particles in the soil, especially the 0.001 mm (0.001 M, 0.001 m) and 0.005 mm (0.005 M, 0.005 m) grain sizes. This makes sense, because fine particles of clay and silt are what drive cohesive behavior. Since this is based on a relatively limited number of cases in the dataset (approximately 100), caution in drawing too strong a conclusion must be taken. However, we can most likely expect that the Atterberg limits (PL and LL) will suffice in serving as a proxy for the amount of fine particles in a soil. This is important since a substantial number of cases do not contain grain size data for the fine particle range, but many have Atterberg limit data.
- Textural measurements from adjacent sieve sizes are highly correlated, as one would expect. The high degree of correlation should allow some of these features to be eliminated without losing too much information. A reasonable approach might be to keep the textural features with the greatest amount of data, like the #4 and #200 sieve sizes. Also, since these are the sieve sizes used to determine USCS soil classification, some estimates of the value for these features can be made if the grain size distribution is not available but the soil class is known.

[†††] Note that an alternative choice might be to eliminate plastic limit, since it is the least correlated of the three plasticity features with CBR. In seeking to retain the most information in the remaining features, I chose to eliminate the one feature correlated most highly with the remaining two.

- Minimum and maximum values of percent passing for each grain size correlate very highly. This is most likely an artifact of the manner in which this data was represented in the database[‡‡‡]. So naturally these cases skew the resulting correlations to indicate a much stronger linear relationship than is actually present. Due this redundancy, an average of the minimum and maximum values for a single sieve size may be a more useful feature than the two individual features themselves.
- Optimum moisture content, maximum dry density, gravimetric moisture content, and field dry density are somewhat correlated with the textural features (especially the #40 and #200 sieve sizes). These four features are also strongly correlated with Atterberg limits and amongst themselves. This could suggest that some of these features may not be needed, however some of the correlation among the four features themselves may be an artifact of the dataset compilation process. Many of the cases reflect several instances of the same soil from a given site, with a single unique test value for maximum dry density and optimum moisture content repeated many times across a group of cases.
- Although depth does not correlate well with any of the other features, it may still be a good candidate for elimination based on intuition. Knowledge regarding the other feature properties that might be a function of depth (e.g., moisture content, grain size, density) should negate its usefulness as a proxy for them.

[‡‡‡] Both were recorded due to the gradation curves for similar soils being grouped together and the plot resulting in a "band" of sizes rather than a distinct curve. However, when a single curve was reported, minimum and maximum values were both recorded as the same value.

With the preceding observations as a guide, four subsets of data (Table 4.3) with increasing numbers of features were selected to focus on in the modeling efforts. The most basic subset of features (A) was selected on the basis of what is expected to be available from the remote sensing platform under the OLS program, namely soil classification and moisture content. With a soil classification, we would hope to know the gradation and plasticity characteristics of the soil, if not directly then at least a range of values for the given soil class. Only cases that had data for each of the features in the subset was used (i.e., no missing data). Also note that to avoid missing values in the plastic limit and plasticity index fields, the data for each subset was separated into plastic and non-plastic groups. The next subset of data (B) includes all the features of the previous plus the dry density parameter, which has been shown to be important in predicting strength. The third subset of data (C) adds the specific gravity feature to those in the previous subset. Specific gravity was included because it is not highly correlated with any other features according to the correlation matrix and can have a large effect on the unit weight of the soil. The final subset (D) added the optimum moisture content and maximum dry density features to all the previous ones. These features were chosen to include after all the others, since they are highly correlated with moisture content and dry density, respectively. With these four subsets of data, the aim was to see how starting with the most important features and then incrementally adding more information might affect the model performance.

Table 4.2: Pairwise Correlation Matrix for Numerical Features* in CBR Database (upper half) and Number of Valid Cases for each Pairing (lower half).

	LL**	PL**	PI**	SpGr	Depth	OMC***	MDD***	MC	DD	3/4 M	3/4 m	#4 M	#4 m	#40 M	#40 m	#200 M	#200 m	0.005 M	0.005 m	0.001 M	0.001 m	CBR
LL**	**1.000**	0.756	0.918	0.550	0.387	0.743	-0.776	0.718	-0.629	0.139	0.286	0.392	0.395	0.477	0.588	0.639	0.629	0.838	0.863	0.887	0.926	-0.409
PL**		**1.000**	0.438	0.557	0.333	0.822	-0.817	0.696	-0.638	0.143	0.251	0.316	0.293	0.460	0.568	0.627	0.574	0.736	0.770	0.815	0.858	-0.225
PI**			**1.000**	0.422	0.334	0.505	-0.560	0.560	-0.466	0.120	0.281	0.355	0.373	0.434	0.533	0.503	0.522	0.817	0.832	0.847	0.876	-0.418
SpGr				**1.000**	0.199	0.421	-0.325	0.315	-0.029	-0.050	0.101	0.104	0.143	0.109	0.233	0.265	0.307	0.119	0.155	0.129	0.149	-0.164
Depth					**1.000**	0.450	-0.458	0.312	-0.453	0.264	0.402	0.442	0.411	0.461	0.490	0.351	0.286	0.314	0.309	0.341	0.337	-0.279
OMC***						**1.000**	-0.918	0.830	-0.746	0.379	0.508	0.527	0.567	0.659	0.754	0.794	0.767	0.722	0.703	0.779	0.740	-0.373
MDD***							**1.000**	-0.760	0.813	-0.551	-0.626	-0.645	-0.653	-0.754	-0.753	-0.708	-0.657	-0.636	-0.625	-0.694	-0.663	0.444
MC								**1.000**	-0.695	0.337	0.456	0.481	0.497	0.591	0.646	0.740	0.694	0.472	0.435	0.516	0.437	-0.436
DD									**1.000**	-0.489	-0.557	-0.684	-0.671	-0.666	-0.621	-0.622	-0.564	-0.428	-0.383	-0.436	-0.357	0.495
3/4 M										**1.000**	0.697	0.798	0.652	0.659	0.512	0.428	0.313	0.336	0.278	0.326	0.267	-0.349
3/4 m											**1.000**	0.891	0.942	0.843	0.797	0.601	0.521	0.421	0.403	0.421	0.394	-0.514
#4 M												**1.000**	0.933	0.907	0.775	0.613	0.541	0.433	0.397	0.435	0.392	-0.536
#4 m													**1.000**	0.916	0.891	0.659	0.646	0.491	0.480	0.502	0.474	-0.578
#40 M														**1.000**	0.901	0.757	0.641	0.532	0.499	0.545	0.499	-0.524
#40 m															**1.000**	0.818	0.794	0.635	0.635	0.655	0.636	-0.507
#200 M																**1.000**	0.951	0.844	0.798	0.853	0.784	-0.448
#200 m																	**1.000**	0.894	0.893	0.893	0.874	-0.447
0.005 M																		**1.000**	0.974	0.989	0.955	-0.313
0.005 m																			**1.000**	0.968	0.990	-0.338
0.001 M																				**1.000**	0.959	-0.355
0.001 m																					**1.000**	-0.331
CBR																						**1.000**

	LL**	PL**	PI**	SpGr	Depth	OMC***	MDD***	MC	DD	3/4 M	3/4 m	#4 M	#4 m	#40 M	#40 m	#200 M	#200 m	0.005 M	0.005 m	0.001 M	0.001 m	CBR
LL**	**1999**	1999	1998	1850	1999	700	705	1612	721	234	234	982	982	243	243	989	989	122	122	104	104	726
PL**		**1999**	1998	1850	1999	700	705	1612	721	234	234	982	982	243	243	989	989	122	122	104	104	726
PI**			**1998**	1849	1998	699	704	1611	720	233	233	981	981	242	242	988	988	121	121	103	103	725
SpGr				**2638**	2638	1117	1128	2192	1216	843	843	1636	1636	843	843	1636	1636	464	464	435	435	1088
Depth					**4592**	1295	1343	4020	1686	1004	1004	1817	1817	1004	1004	1838	1834	496	496	466	466	1533
OMC***						**1295**	1287	1127	813	681	681	1197	1197	678	678	1207	1207	365	365	340	340	698
MDD***							**1343**	1175	861	686	686	1217	1217	683	683	1227	1227	368	368	343	343	733
MC								**4020**	1656	880	880	1583	1583	876	876	1601	1600	410	410	388	388	1476
DD									**1686**	683	683	997	997	678	678	1011	1010	346	346	339	339	1380
3/4 M										**1004**	1004	995	995	993	993	1004	1004	496	496	466	466	526
3/4 m											**1004**	995	995	993	993	1004	1004	496	496	466	466	526
#4 M												**1817**	1817	1003	1003	1817	1817	496	496	466	466	849
#4 m													**1817**	1003	1003	1817	1817	496	496	466	466	849
#40 M														**1004**	1004	1004	1004	496	496	466	466	526
#40 m															**1004**	1004	1004	496	496	466	466	526
#200 M																**1838**	1834	496	496	466	466	863
#200 m																	**1834**	496	496	466	466	861
0.005 M																		**496**	496	466	466	269
0.005 m																			**496**	466	466	269
0.001 M																				**466**	466	257
0.001 m																					**466**	257
CBR																						**1533**

***** Key to abbreviations and acronyms used for features can be found in Appendix A.

****** Atterberg limits for cohesive soils only.

******* Standard CE-55 compaction [17].

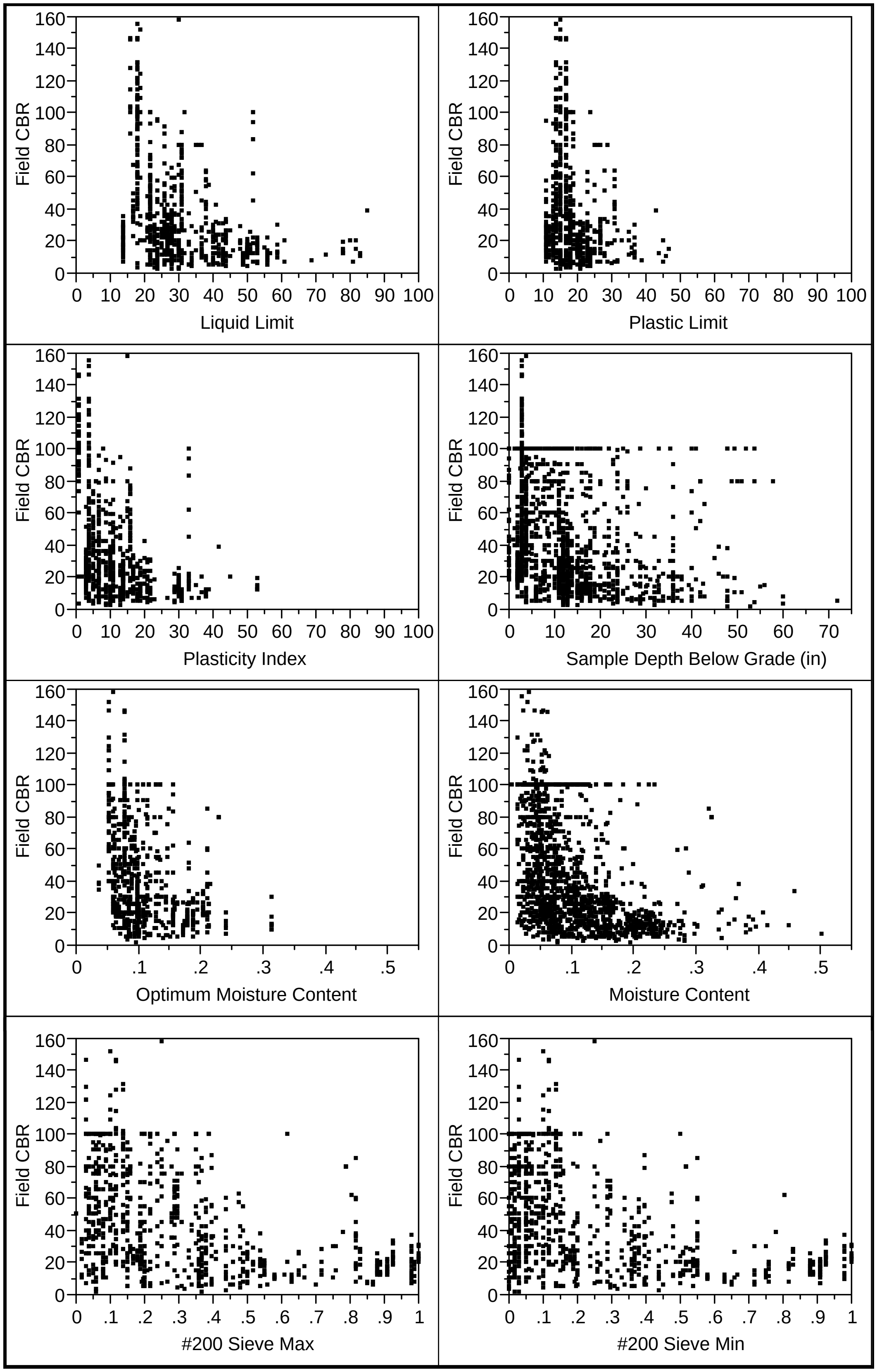

Figure 4.5: Pairwise Scatterplots of Numerical Features with an Inverse Relationship to CBR.

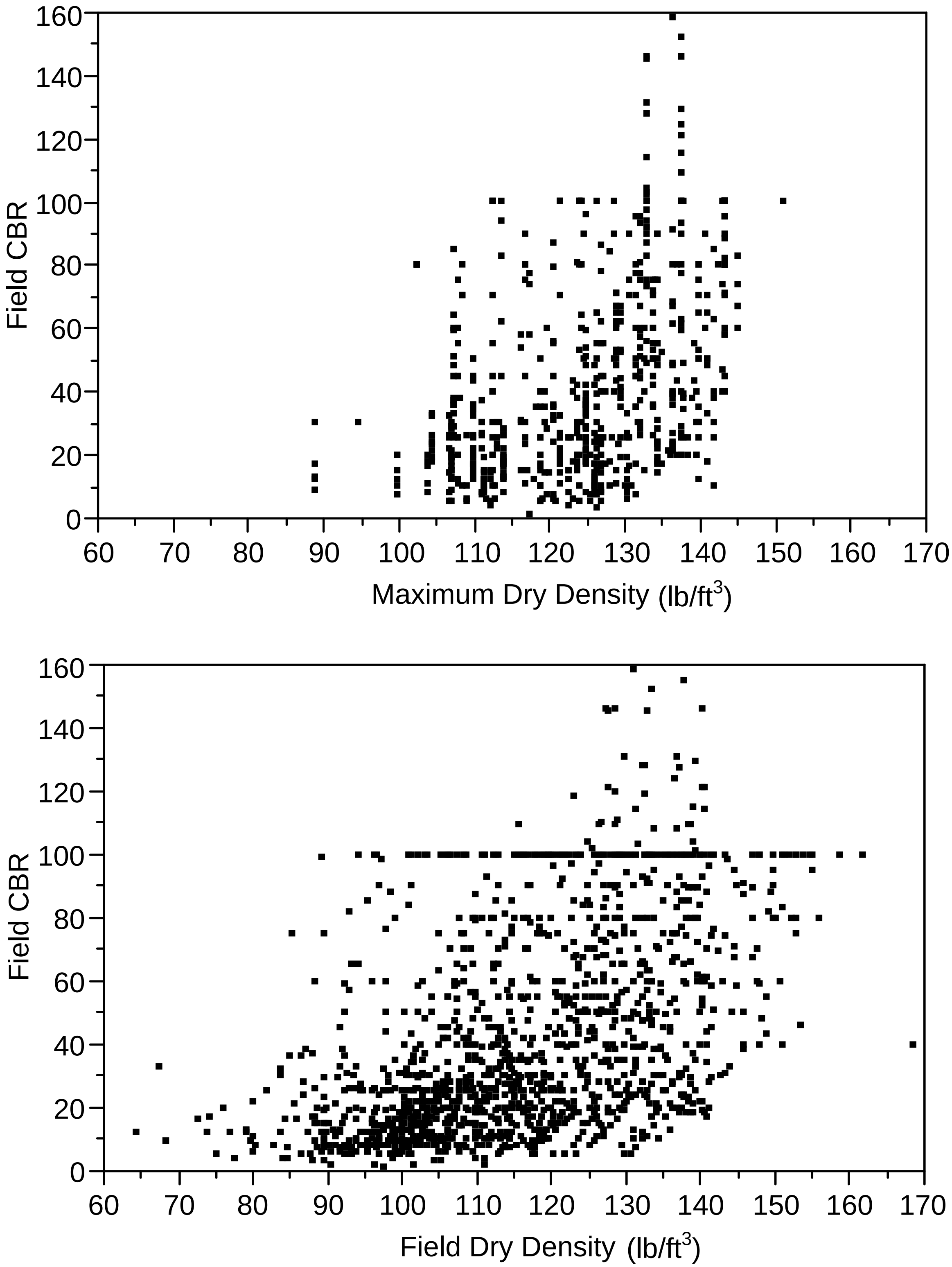

Figure 4.6: Pairwise Scatterplots of Numerical Features with an Exponential Relationship to CBR.

Table 4.3: Subsets of OLS CBR Database with Reduced Numbers of Features used for Model Analysis.

Feature*	Subset A		Subset B		Subset C		Subset D	
	Plastic	Non-Plastic	Plastic	Non-Plastic	Plastic	Non-Plastic	Plastic	Non-Plastic
PL**	X		X		X		X	
PI**	X		X		X		X	
#4 Avg	X	X	X	X	X	X	X	X
#200 Avg	X	X	X	X	X	X	X	X
MC	X	X	X	X	X	X	X	X
DD			X	X	X	X	X	X
SpGr					X	X	X	X
OMC***							X	X
MDD***							X	X
# of Cases	363	454	341	430	311	363	294	251

X Indicates that feature is included in the subset of data.
* Key to abbreviations and acronyms used for features can be found in Appendix A.
** Atterberg limits for cohesive soils only.
*** Standard CE-55 compaction [17].

4.4 Data Pre-processing

Prior to using the four data subsets with some of the prediction methods, normalization and transformation of the data were necessary because the models perform best with properly scaled data. Several approaches are available to normalizing the data so that each feature is roughly the same magnitude as all the others and the scale differences do not bias results. A minimum-maximum normalization, where the data is rescaled so that the minimum is -1 and the maximum is 1, is available in the MATLAB Neural Network Toolbox [114]. A second normalization, where the data is rescaled to a mean of zero and unity standard deviation, also exists. A third normalization strategy was custom-written for this study based on the MATLAB mean-standard deviation code. In this case the data is rescaled to have a median of zero and unity interquartile range (25^{th} percentile to the 75^{th} percentile of the data). It has been suggested that this third

technique may be less susceptible to the effect of outliers in the data [98]. In testing with the machine learning methods, all three normalization techniques were tried and provided equivalent results. The second normalization method of scaling to zero mean and unity standard deviation was chosen for use throughout testing. This technique is commonly used and does not present scaling problems in practice if new inputs to the model fall somewhat outside the minimum-maximum range of the original training dataset.

Normalization is required for various reasons with each of the prediction methods, as discussed previously (section 4.2). The *k*-nearest neighbor method is sensitive to scaling of features, so the inputs were normalized for this method. Target CBR values were not scaled for this method, since they do not affect the "learning" process of determining similarity among neighbors. For artificial neural networks, both the input features and the targets were normalized to avoid problems with the network weights from becoming oversaturated when trying to adjust for the sigmoid activation functions that only range from -1 to 1. For trials with multivariate linear regression, both the inputs and output CBR were not normalized.

For the artificial neural network and multivariate linear regression methods, a second pre-processing step was performed on the target CBR values. In order to treat the size of errors at the low and the high ends of the scale more equally, a logarithmic transformation of the target values was used. The motivation behind this lies in the fact that these two prediction methods are based on a least squares approach to minimizing the mean error over the whole training set. Without a transformation that makes errors proportional at all points along the scale, the errors at the higher end of the scale will dominate the mean error term. In this case, the optimization process will mostly

concentrate on reducing errors at the high end, while ignoring sizable percentage errors at the low end. In order to avoid this, a base 10 logarithm was used to transform the CBR targets. Although it does not perfectly equalize the lower and upper ends of the scale, it makes a large improvement over the unscaled approach. For example, the same percentage error at a CBR of 100 and a CBR of 10 would compare with each other by a factor of two in the transformed state, but by a full order of magnitude if untransformed. This approach to data transformation is always recommended for data that exhibit a multiplicative response, where local variation is proportional to its absolute value [98].

The justification for a logarithmic transformation of the target CBR values lies in the fact that errors in strength measurement for weak materials are considered more critical. The military standard for conducting the field CBR test provides guidance on what is considered a "reasonable agreement" among three test values at one location, listed in Table 4.4 [14]. If the three tests do not fall within these ranges, then three more measurements are made and grouped together with the original ones. The numerical average of the six measurements are then used to give a more representative result. From the table, it is clear that on an absolute basis the precision of the field CBR measurement is most important at the lower end of the scale.

Table 4.4: DoD Guidance on the Permissible Range of Three Field CBR Measurements.

CBR Range (%)	***Permissible Tolerance (in CBR units)***
Less than 10	3
10 to 30	5
30 to 60	10
Above 60	Unimportant

4.5 Estimating Generalization Error

For any prediction problem, generalization of the prediction model to new cases is clearly an important issue. As discussed earlier (section 4.2.3), models tend to overspecialize in that they perform better on the cases that they have been trained with, but not as well on unseen cases. In order to accurately estimate the error rate for new data, performance on a "test set" of data that has been set aside during the model-building process must be used. The difficulty of choosing this test set so it is a representative sample of the dataset and not biased in some way must be addressed.

A common technique for estimating the error on unseen cases is known as cross-validation, which is based on a repeated "resampling" of the entire dataset [102]. In "v-fold cross-validation", the dataset is split up into v random subsets of approximately equal size. Then, using v-1 of the subsets, a model is built and the error is estimated with the one subset that was left out. The process is repeated v times, each time leaving out a different one of the subsets, which results in v separate models. The v independent test set errors are then averaged together, to come up with an overall estimate for the generalization error. In this way, the variability in the error estimate from randomly choosing different test sets is spread out over the v trials and a more stable average value results. Ten is a popular choice for v when estimating generalization error [13].

A logical extension of v-fold cross-validation is choosing v so that it the same size as the number of cases in the entire dataset. This is known as "leave-one-out" cross-validation which is robust, but can be very time intensive when used with models that are computationally expensive to train and build. Comparisons between leave-one-out and 10-fold cross-validation during the k-nearest neighbor testing in this

study revealed close agreement in the resulting error estimates. Therefore, 10-fold cross-validation was used with confidence throughout this investigation for comparison of performance among all models tested.

The generalization error is often used to choose among different models and assess the choice of parameters that lead to the best performance [13]. The error metric chosen for use in this study was the Normalized Root Mean Square Error (*NRMSE*). This is calculated in the following manner:

$$NRMSE = \left[\frac{\sqrt{\dfrac{\sum_{1}^{N} (y_i - \hat{y}_i)^2}{N}}}{\bar{y}} \right] \times 100 = CV \tag{4.4}$$

where:

N	=	Number of cases in the dataset
y_i	=	Known target value for the i^{th} case in the dataset
$\hat{y}_i$	=	Prediction for the i^{th} case in the dataset
$\bar{y}$	=	Mean of the known target values in the dataset

Essentially it is the standard deviation of the error residuals normalized relative to the mean of the targets. It can be readily compared across different trials in a relative sense, independent of the dataset or model type. It is also equal to the coefficient of variation (*CV*), which is a measure of the precision of the test method (see discussion in section 5.2.1).

5 Results

5.1 Prediction Method Trials

In all the rounds of testing carried out, several procedures remained constant throughout. All models were implemented using the MATLAB software package [121]. Unless otherwise noted, 10-fold cross-validation was used in all trials when estimating error and choosing model parameters. Error rates were calculated by taking the average across all validation folds. When comparing model performance on one data subset, the same exact split of the data into the cross-validation folds was used for all methods. In this way, any minor variability in the error due to different sorting of the data was eliminated, which allowed a more direct comparison. For the data pre-processing tasks of normalization and transformation, these were carried out independently on the training, validation, and test sets of the data. By preprocessing these separately for each fold of the cross-validation, any biasing of the test set data with information about the training data was avoided.

5.1.1 *k*-Nearest Neighbor

In order to apply the *k*-nearest neighbor method for a regression problem, a customization of some available MATLAB code was necessary. The Bioinformatics Toolbox contains a function named *knnclassify*, which implements the *k*-nearest neighbor method to perform a classification task. This formed the basis for a custom-written function that operates in a similar way but performs regression. Three different distance metrics were evaluated to determine similarity among cases: Euclidean, city-block, and cosine. Also, three weighting schemes for generating a prediction from the resulting target values were programmed: a straight average of the *k* neighbors, the inverse of the

squared distance, and the exponential of the negative distance. (See section 4.2.2 for an explanation of these distance and weighting techniques.)

To select the *k* value that resulted in the best performance, 10-fold cross-validation was carried out while letting *k* range from one to twenty. The average normalized root mean square error on the test set over the ten folds was calculated at each value of *k*. This was performed on feature subsets A, B, C, and D for both plastic and non-plastic soils at each combination of the three distance metrics and three weighting schemes. Performance was generally comparable across all nine, so the simplest approach (Euclidean distance and average weighting) was chosen for further testing. For plastic and non-plastic soils, different *k* values worked best at providing the best normalized root mean square error on the test set. A *k* value of four worked best for plastic soils, and a *k* of ten worked well for non-plastic soils.

Using these parameters for the number of neighbors, the method was tested on each of the A, B, C, and D feature subsets for both plastic and non-plastic soils. In each case the inputs were normalized to zero mean and unity standard deviation, due to the method's sensitivity to the relative scaling of the features. The target CBR values were not pre-processed. The normalized root mean square error results on the test set data for the k-nearest neighbor models are shown in the first row of Table 5.4.

5.1.2 Artificial Neural Networks

For all trials involving artificial neural networks, the use of several procedures were common to all testing throughout this study. Target values were transformed with a base 10 logarithm. Subsequently, both input and target variables were normalized to zero mean and unity standard deviation prior to use with the ANNs. Neural network

architecture included input nodes equal to the number of inputs in each dataset, one hidden layer of neurons[§§§], and one output neuron for the CBR target. All networks were feed-forward back-propagation neural networks[****]. A tangent sigmoid activation function[††††], ranging from 1 to -1, was used for the hidden layer neurons to provide the nonlinear flexibility for the network. For the output layer neuron, a linear function was used as the activation function to allow the mapping of the target value in a linear fashion. When running the ANN models, four independent runs of the network with randomized values for the initial weights were made for each condition (different architectures, learning methods, etc.). This was done to help identify and minimize any problems with the network converging to local minima.

Different combinations of learning methods were used during separate rounds of neural network testing, however within each round these were kept the same. For all learning methods employing the early stopping technique to avoid overfitting of the data, the subset of data was randomly split into three groups. Seventy percent of the data was placed in the training set, twenty percent was used for the validation set, and ten percent was set aside for the test set. The ten percent allocated to the test set was simply the holdout fold for the 10-fold cross-validation. One regularization method, Automated Bayesian Regularization, was also tested. For this method the same seventy percent of the data was used for the training set and the same ten percent holdout as the test set.

[§§§] Some trials with two hidden layer were run, however no improvement in performance resulted. The lowest generalization error on the test set always occurred with one hidden layer.
[****] A few limited tests with generalized regression neural network using radial basis functions were carried out, however there was no improvement in performance over the feed-forward back-propagation networks.
[††††] In addition to the tangent sigmoid and radial basis functions, a log sigmoid function was also tested but no improvement in performance was observed.

5.1.2.1 Four Feature Subsets

Artificial neural network testing was carried out on each of the A, B, C, and D feature subsets for both plastic and non-plastic soils. Three of the MATLAB implementations of the early-stopping learning methods for updating the network weights were used in the testing. Each one represented a different approach to error minimization: steepest gradient with adaptive learning rate & momentum, Powell-Beale conjugate gradient, and Levenberg-Marquardt (a Newton-type method). Performance among the three approaches was comparable. The optimal number of hidden layer neurons was determined using a 10-fold cross-validation over a range of values. These were bounded by letting the total number of weights and biases vary between one-half to one sixty-fourth of the number of training cases in each dataset. Typically the best performance resulted with about 8 neurons for most of the datasets. The lowest normalized root mean square error results for the test set data are shown in the second row of Table 5.4.

5.1.2.2 Single Features

In an attempt to get a general sense of the "nonlinear correlation" of each of the numerical features in the database with the CBR target field, a series of neural networks were run with single features as the only input. By assessing the lowest error rate that could be achieved for each, it was hoped that this might give some additional insight into which features were the most important for predicting CBR. The number of neurons in the hidden layer were varied from three to fifteen, by steps of three. For learning methods, eight of the early-stopping approaches available in MATLAB were run plus automated Bayesian regularization. Performance among the learning method approaches

was similar and the reported error rates represent the lowest from the group. The number of cases with data available for each feature varied widely, from a low of 257 for the percent passing the 0.001 mm grain size to 1,533 for the depth below grade level. For all other features, the number of cases can be seen in the right-hand column in the lower half of Table 4.2. Because the number of cases varied so widely from feature to feature or possibly the degree of nonlinearity was different in their relationships to CBR, the number of neurons producing the lowest error rate ranged from three to twelve in the trials. The lowest normalized root mean square error on the test set data for each feature is provided in Table 5.1 in ascending order. As a comparison, the linear correlation coefficients (R) between each of the features and CBR are included in descending order on an absolute basis.

Table 5.1: Comparison of Single Input ANN Performance and Correlation Coefficients for Each Feature.

Feature*	**NRMSE for Single Input ANN to Predict CBR**
3/4 m	55.4
#40 M	58.2
#40 m	58.8
3/4 M	58.9
#4 m	59.8
#4 M	60.3
#200 M	63.9
#200 m	65.3
MDD******	67.9
DD	69.3
OMC******	69.6
MC	69.9
0.001 m	70.7
LL**	72.1
0.005 m	72.2
0.001 M	73.3
0.005 M	73.6
Depth****	73.9
SpGr	75.8
PI**	79.8
PL**	87.6

Feature*	***Pearson Product-Moment Correlation (R) with CBR***
CBR	1.000
#4 m	-0.578
#4 M	-0.536
#40 M	-0.524
3/4 m	-0.514
#40 m	-0.507
DD	0.495
#200 M	-0.448
#200 m	-0.447
MDD******	0.444
MC	-0.436
PI**	-0.418
LL**	-0.409
OMC******	-0.373
0.001 M	-0.355
3/4 M	-0.349
0.005 m	-0.338
0.001 m	-0.331
0.005 M	-0.313
Depth****	-0.279
PL**	-0.225
SpGr	-0.164

***** Key to abbreviations and acronyms used for features can be found in Appendix A.
****** Atterberg limits for cohesive soils only.
******* Depth below grade level.
******** Standard CE-55 compaction [17].

5.1.2.3 Subsets of the Highest Correlated Features

Concentrating on the features that had the best performance individually, several neural networks were tested using only these inputs. Results for both the linear correlation of the features to CBR and the single input ANNs showed a similar grouping of top features (Table 5.1). The coarser particle sizes (¾ inch, #4 & #40 sieves)

produced the six leading features for the single input ANNs, and five of these six features were also the strongest linear correlations. A subset of the data, containing entries for all six features associated with these grain sizes, was selected from the full database and split into plastic and non-plastic groups. These datasets were used to train and test two neural networks and the normalized root mean square error for the test set from cross-validation is shown at the left side of Table 5.2. In addition, the top four features for both the linear correlation ranking and the single feature input ANNs were tested separately as shown in the remaining columns of Table 5.2. The database cases used for testing were identical throughout the three trials with different input combinations. The three learning methods and the approach to bounding the number of hidden neurons were the same as those described previously for the four main feature subsets (section 5.1.2.1). Performance among the learning method approaches was similar and the reported error rates represent the lowest of the three.

Table 5.2: Comparison of ANN Performance on Highest Correlated Features.

Input Features*	***Six Best Overall***		***Four Best Linear Correlation***		***Four Best ANN NRMSE***	
	Plastic	*Non-Plastic*	*Plastic*	*Non-Plastic*	*Plastic*	*Non-Plastic*
3/4 m	X	X	X	X	X	X
3/4 M	X	X			X	X
#4 m	X	X	X	X		
#4 M	X	X	X	X		
#40 m	X	X			X	X
#40 M	X	X	X	X	X	X
# of Cases	92	420	92	420	92	420
NRMSE	68.7	51.4	75.4	51.0	70.8	51.2

X Indicates that feature is included in model input.
* Key to abbreviations and acronyms used for features can be found in Appendix A.

5.1.2.4 Single Soil from One Site

To determine whether lower generalization error rates might be possible using a more targeted approach, a single USCS soil type from one geographic location was chosen for further testing. A low-plasticity clay (CL) from the Santa Fe Municipal Airport site was selected because it is one of the most common USCS soils (see Figure 3.5) and represented the greatest number of cases (42) for a single soil type at one place in the database. Different combinations of features were used in neural network trials, with four of these matching the feature sets for the four main datasets (A, B, C, & D) tested in section 5.1.2.1. In addition a fifth combination of features using only those that contained data for most cases was also tried. The average normalized root mean square error results on the test set from cross-validation are shown in Table 5.3 for all these feature combinations. All nine learning methods used for the single input ANNs (section 5.1.2.2) were tested and the number of hidden neurons varied from one to three. Performance of each learning method was similar and the reported error rates represent the best among the group.

Table 5.3: Performance of an Artificial Neural Network on a Single USCS Soil Type from a Single Site (CL soil from Santa Fe Municipal Airport).

Input Features*	***Subset A Inputs***	***Subset B Inputs***	***Subset C Inputs***	***Subset D Inputs***	***Most Complete Fields***
PL	X	X	X	X	X
PI	X	X	X	X	X
#4 Avg	X	X	X	X	
#200 Avg	X	X	X	X	
MC	X	X	X	X	X
DD		X	X	X	X
SpGr			X	X	X
OMC* ***				X	
MDD* ***				X	
# of Cases	20	20	20	19	41
NRMSE	49.8	43.8	46.5	55.9	51.1

X Indicates that feature is included in model input.
***** Key to abbreviations and acronyms used for features can be found in Appendix A.
******* Standard CE-55 compaction [17].

5.1.3 Multivariate Linear Regression

Multivariate linear regression was carried out on feature subsets A, B, C, and D as a comparison for the *k*-nearest neighbor and artificial neural network performance. The target CBR values were first transformed using a base 10 logarithm, but left unnormalized. The input features were not pre-processed. The results for the linear regression models are shown in the third row of Table 5.4.

Table 5.4: Comparison of Average NRMSE Performance of Test Set Predictions on Four Feature Subsets with 10-fold Cross-Validation.

Prediction Method	***Subset A***		***Subset B***		***Subset C***		***Subset D***	
	Plastic	*Non-Plastic*	*Plastic*	*Non-Plastic*	*Plastic*	*Non-Plastic*	*Plastic*	*Non-Plastic*
k-NN	54.7	57.4	53.0	49.1	48.2	45.6	54.3	53.5
ANN	58.7	51.8	54.6	50.9	56.4	48.1	54.9	50.5
Linear Regression	63.2	53.6	62.6	51.8	63.7	48.6	59.3	50.8
Naïve Mean	88.0	56.3	89.5	55.3	91.1	53.6	88.6	54.3

5.2 Performance Baselines & Comparison Measurements

5.2.1 Upper and Lower Limits on Performance

In order to get a sense of the relative performance of the models, constraints on the best and worst error rates that might be expected are useful. One metric that can be used as a baseline to represent a very poor prediction method is the concept of a “naïve mean” model [122,123]. This idea simply involves using the average value of all the target CBR values in the training set as a prediction for the cases in the test set. In effect, all the input features are ignored when making new predictions. This method was applied with 10-fold cross-validation to the four feature subsets of data for both plastic and non-plastic soils. The test set performance results for each are given in the fourth row of Table 5.4. In addition, the same approach was used with the subset of all 1,533 cases that contained all numeric CBR target values in the database and the normalized root mean square error was 76.8 percent.

While the naïve mean model represents only a poor model and not necessarily the worst possible, there are more precise methods for determining the performance limit for a perfect model. The fundamental threshold for prediction accuracy is tied to the variability of the response parameter itself. "If you have noise in the target values, the mean squared generalization error can never be less than the variance of the noise, no matter how much training data you have" [13]. The minimum error rate that a perfect predictor can achieve is known as the "Bayes error rate" [102,124].

To assess how "noisy" the CBR test method is, reports of test precision were gathered from the literature. In particular, two reports contained a wealth of information on the variability of pavement materials, including many references for CBR [93,125]. The coefficient of variation, which represents the precision of a group of measurements, are given in Table 5.5 for lab and field CBR tests over a range of strength values. These represent the variability encountered within uniform lots of material from a jobsite.

Some trends are apparent in the data. Higher strength materials generally exhibit more variability, which would be expected based on the limitations of the CBR test with granular materials. Field measurements appear to have greater variability, as anticipated. The natural soils and the more "select" materials used in pavement subbase and base layers seem to have comparable levels of variability. Less variability might be expected in the select materials, so the limitations of the CBR method for granular materials may also be a factor here.

The coefficient of variation ranges from about 6 to 38 percent overall, but most sources commonly fall somewhere near 25 percent. Consequently, it is reasonable to

conclude that the normalized root mean square error for even a perfect prediction model will be limited to this level of generalization error.

Table 5.5: Reported Variability of CBR Measurements in the Literature (after [93,125]).

Pavement Layer or Soil Deposit	***Mean CBR (%)***	***Standard Deviation (%)***	***Coefficient of Variation (%)***	***Number of Tests***	***Lab or Field CBR Test***	***Reference***
Subgrade	7.8	1.4	17.9	7	Field	[28]
Subgrade	N/A	N/A	20	N/A	Field	[128]
Subgrade	N/A	N/A	20	N/A	Field	[132]
Subgrade	4.2	N/A	21.4	33	Field	[28]
Subgrade	7.1	1.5	22.3	33	Field	[28]
Subgrade	18.2	4.7	26.2	7	Field	[28]
Residual Fine-Grained Soil Deposit	10	2.7	27	N/A	N/A	[126]
Subgrade	1.14	N/A	27.5	18	Field	[127]
Subgrade	1.445	N/A	27.5	11	Field	[127]
Subgrade	N/A	N/A	35	8 lots	Field	[128]
Engineered Fill	21	6.7	32	N/A	N/A	[129]
Engineered Fill	12*	4.1*	34*	N/A	N/A	[129]
Engineered Fill	43	15	35	N/A	N/A	[130]
Subbase	N/A	N/A	8.6	N/A	Field	[128]
Subbase	N/A	N/A	20	N/A	Field	[132]
Subbase	N/A	N/A	22	12 lots	Field	[128]
Subbase	26.3	8.3	31.9	33	Field	[28]
Subbase	20.3	N/A	36.9	33	Field	[28]
Subbase and Base	59	13	22	N/A	Field	[130]
Subbase and Base	N/A	N/A	7 to 26	N/A	Lab	[131]
Base	N/A	N/A	20	N/A	Field	[132]
Base	94.3	36	37.6	72	Field	[28]
N/A	74	4.4	5.9	7	Lab	[15]
N/A	17.2	1.41	8.2	7	Lab	[15]

* Estimated by dynamic cone penetrometer (DCP) tests.

5.2.2 Performance of Existing “Universal” Methods

As another attempt to assess the relative performance of the models developed in this study, some of the existing CBR prediction methods found in the literature were evaluated with the OLS database. Because these models were not built upon the OLS CBR database, the entire dataset could be used to evaluate them without using the cross-validation approach.

The method of estimating CBR with typical ranges of values depending on the Unified soil classification was tested on the OLS data (detailed method description in section 2.2.1). Specifically, for each USCS soil type, the average of the CBR range for that soil class was used as a prediction. This approach was tried for each of the five separate ranges of values cited in the literature. The prediction error for all 1,533 cases with CBR values is presented in Table 5.6, as well as a breakdown by plastic and non-plastic soils, and by each individual USCS soil type.

Another way to look at the typical CBR ranges for USCS soil types is as a classification method. In comparing the typical ranges to the distribution of the soil strengths for each soil type in the full dataset, the approach was not very successful. For the gravel soils (GW, GP, GM, and GC), the typical CBR ranges reported in the literature overlapped closely with the interquartile ranges for each soil type in the database. However, this represents only 50 percent of the data, so the approach does not capture the other half of the cases—even for a relatively predictable material like gravel. The CBR ranges for other soil types did not show this relationship with the interquartile range, and in general captured an even lower proportion of the cases in the full dataset.

Table 5.6: Normalized Root Mean Square Error for using the Average of Typical Unified Soil Classification System California Bearing Ratio Ranges (from Table 2.2) as a Prediction.

USCS Soil Type	*USACE [30], US Army [31], and US Army & Air Force [27]*	*Yoder & Witczak [28]*	*US Army, Air Force & Navy [14] and PCA [29]*	*Rollings & Rollings [18]*	*NCHRP [32]*
Overall	84.2	77.8	81.2	78.0	75.7
Plastic Soils Only	97.9	90.2	92.7	90.2	91.6
Non-Plastic Soils Only	67.1	61.0	65.6	61.1	61.9
GW	45.8	49.5	49.5	49.5	49.5
GP	56.3	55.9	57.2	55.9	55.9
GM	64.5	47.6	55.3	47.6	51.3
GC	66.4	66.4	66.4	66.4	68.6
SW	53.6	53.6	53.6	48.3	53.6
SP	82.5	88.9	92.6	92.6	85.6
SM	75.4	70.1	75.4	70.1	70.1
SC	107.7	103.9	103.9	103.9	103.9
ML	107.3	101.7	101.7	101.7	97.5
CL	76.7	66.8	66.8	66.8	66.8
OL	--	--	--	--	--
MH	83.4	78.1	78.1	78.1	83.4
CH	113.9	125.5	125.5	125.5	129.3
OH	97.5*	96.0*	96.0*	96.0*	--
Pt	--	--	--	--	--
CL-ML	--	--	--	--	--
GW-GM	--	--	--	--	45.2
GW-GC	--	--	--	--	73.9*
GP-GM	--	--	--	--	55.8
GP-GC	--	--	--	--	16.7*
GC-GM	--	--	--	--	--
SW-SM	--	--	--	--	82.4
SW-SC	--	--	--	--	--
SP-SM	--	--	--	--	85.4
SP-SC	--	--	--	--	--
SC-SM	--	--	--	--	--

* Marked results are based on a single case in the database. Caution is needed for drawing conclusions for these soil types. See Table 3.3 for the number of cases for other soil types.

-- No cases in the database or no range given in the literature to make a prediction with.

A second existing method, from the Mechanistic Empirical Design Guide [32], was also tested on some of the cases in the database. The method, outlined in section 2.2.1 uses different approaches for plastic and non-plastic soils. For plastic soils, a total of 336 cases in the database that contained information on both the plasticity index and percent passing the number 200 sieve size (average) were used. For non-plastic soils, the method requires the D_{60} parameter, which is the diameter on the cumulative size distribution curve where 60 percent of particles are finer. For 433 cases in the database where the gradation data contained the 60th percentile for non-plastic soils, a linear interpolation between the two closest sieve sizes was used. Prediction results for the method on both soil groups are presented in Table 5.7.

In order to assess whether artificial neural networks could provide any improvement given similar data, several models were run alongside the Mechanistic-Empirical Design Guide method. For the plastic soils, the same dataset and inputs were used for the ANN. In the case of the non-plastic soils, two separate attempts with somewhat reduced datasets were tried. Fewer cases were used for these because the ANNs were limited to cases with complete data for all input features. One model used four of the coarser sieve size features as inputs, while a second included two more that measured the fine particles. The model with fine particles was tried separately, due to the lack of data for about half of the cases and the expectation that non-plastic soil behavior should not be related to fines very strongly. Generalization error for these ANNs are provided in Table 5.7 alongside the Mechanistic-Empirical Design Guide results for comparison. The same three learning methods and the bounds on the number of weights

and biases for the ANNs were used as the tests on the four feature subsets described in 5.1.2.1.

Table 5.7: Performance of Mechanistic-Empirical Design Guide method versus Artificial Neural Networks given Comparable Input Features.

Input Features*	***Plastic Soils***		***Non-Plastic Soils***		
	M-E Design Guide	*ANN*	*M-E Design Guide*	*ANN*	*ANN*
PI	X	X			
3/4 Avg			X	X	X
#4 Avg			X	X	X
#40 Avg			X	X	X
#200 Avg	X	X	X	X	X
0.005 Avg			X		X
0.001 Avg			X		X
# of Cases	336	336	433	400	212
NRMSE	79.1	74.8	51.9	49.9	47.3

X Indicates that feature is included in model input.
***** Key to abbreviations and acronyms used for features can be found in Appendix A.

5.3 Assessing the Reliability of Predictions

In order to provide some indication of the confidence that can be associated with a specific prediction, the distribution of the error residuals for the test set data can be used to estimate confidence intervals for the model predictions. By doing this, we are assuming that the estimated generalization NRMSE is a good approximation of the variability that we would expect to encounter with new data that the model has not seen before. For all the prediction methods tested in this study, the distribution of error residuals was normal. Thus, we can use the normalized root mean square error as a direct, normalized estimate of the standard deviation of the error for new predictions.

But, what level of confidence is needed for soil strength when applying traffic loads? For fixed facilities that are constructed by DoD, there is some statistical guidance

for selecting a design CBR from a group of field tests that exhibit the variability of soil strength measurement [133]. For these airfields, built to have a design life of twenty years, the CBR is selected from among all the measurement values so that eighty-five percent of the readings equal or exceed the design value. In effect, the design CBR is chosen to be the 15^{th} percentile of the field CBR measurements.

Presumably for contingency operations, like the scenarios that would be supported by the OLS approach, this degree of confidence would change. But, whether the confidence levels would shift lower or higher is not clear. At first, it would seem that in a crisis that greater levels of risk would be acceptable, so confidence levels of less than eighty-five percent might be adequate. However, landing on an area where there has been no prior ground reconnaissance would seem to introduce many unknowns into the risk assessment, making confidence levels of greater than eighty-five percent on the soil strength necessary. Ultimately, it is up to the military leaders, mission planners, and aircraft pilots to make these types of decisions. But in the meantime, proceeding with the current level of confidence used for picking strength from a set of direct measurements is useful for illustration.

For a normal distribution, one standard deviation below the mean represents the 15.87 percentile of the data. So, we can subtract the normalized root mean square error from predictions to provide approximately the same level of confidence as for field in-place design CBR values used for fixed facilities. This approach to reducing the predicted value by using the NRMSE is illustrated in Figure 5.1, with three levels of error to demonstrate how the error magnitude affects the process. For models with a NRMSE of fifty percent, typical of the best performers found in this study, this means that model

outputs must be cut in half to provide predictions where eighty-four percent of the time we could expect the real soil strength to meet or exceed this. If we could achieve an error rate closer to the Bayes error limit posed by the variability of the CBR measurement itself, the reduction would only need to be one quarter of the model output.

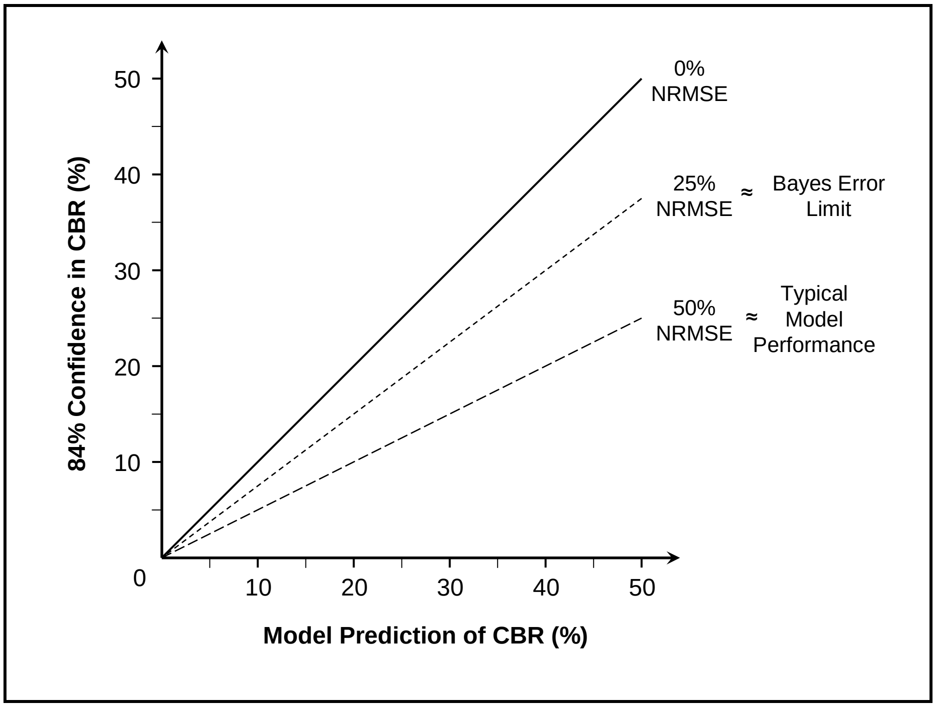

Figure 5.1: The Effect of the Normalized Root Mean Squared Error Magnitude on Prediction Confidence.

Some actual examples of how this need to overestimate strength levels in order to provide an acceptable level of confidence are provided in Table 5.8. In order to support ten passes of C-130 H and C-17 aircraft on an unsurfaced airfield, the soil surface strength requirements are shown [20][‡‡‡‡]. Assuming an NRMSE of fifty percent, the model predictions would need to be double the necessary strength in order to provide the necessary confidence.

[‡‡‡‡] Note that the strengths provided are only the minimum necessary for the soil surface layer of the airfield. Other levels in the pavement structure must be evaluated separately to determine the critical layer that will determine the overall strength of the system as a whole.

Table 5.8: An Example of Soil Surface Strength Requirements for Unsurfaced Operations of Cargo Aircraft and the Effect of 50% NRMSE on Prediction Confidence.

Aircraft	Gross Weight (lbs)	Surface CBR Required for 10 Passes	Model Prediction Needed for 84% Confidence
C-130 H			
Minimum weight	69,000	2	4
Maximum weight	175,000	6.5	13
C-17			
Minimum weight	279,000	4	8
Maximum weight	585,000	12.5	25

6 Discussion

Taken as a whole, the results demonstrate that the use of machine learning methods can significantly improve strength prediction accuracy for soils that exhibit plastic behavior but provide no benefit for non-plastic soils. In comparison to the naïve mean baseline error rate, the advanced techniques are able to halve the error rate for plastic soils. On an absolute basis however, the performance of these plastic soil models still only provide normalized root mean square error rates of about fifty percent—the same meager performance as for the non-plastic soil. For non-plastic soils, none of the prediction methods tested appear to provide any significant improvement over the naïve mean baseline.

Based on the baseline naïve mean results (Table 5.4) it is clear that plastic soils exhibit greater variability than non-plastic soils. This is expected, because plastic soils are more susceptible to considerable strength changes with relatively minor fluctuations in moisture content. This nonlinear behavior lends itself well to being modeled by the machine learning methods, with *k*-nearest neighbor providing the best performance.

The increasing number of features incorporated into the models when progressing from data subsets A through D (Table 5.4) provide some insights regarding the importance of these inputs. For both *k*-NN and ANN models, adding dry density to the plasticity, moisture content, and particle size features leads to better performance. Subsequently, adding specific gravity yields the best overall results for the *k*-nearest neighbor method, but does not benefit the artificial neural network model. In fact, the neural network performance did not change very much across all four feature subsets for the generalized models. For both *k*-NN and ANN methods, it appears that the Proctor

relationships among optimum moisture content and maximum dry density are less important if all the previous features are included, which is very good since these values would most likely be very difficult to determine from remotely-sensed data. In terms of feature selection , it appears that an intuitive choice of inputs based on knowledge of the problem domain actually works best. In fact, for plastic soils an intuitive choice of features provided roughly twice the reduction in baseline error rate versus choosing features based simply on relative correlation with strength (Table 5.2).

The use of single feature artificial neural networks to determine the relative importance of features presented results with remarkable similarity to the order determined using the pairwise linear correlation approach, as shown in Table 5.1. Both revealed that coarse particle sizes are most important in determining strength, which is not surprising. Materials with coarse grains are most likely to exhibit a consistent strength, being less susceptible to changes due to moisture content and density. Therefore, one would expect that the presence or absence of large grain sizes could be most easily associated with strength on an individual basis.

Testing the performance of two existing methods versus the cases in the OLS database revealed that they do not work very well at all. Using the average of typical CBR ranges based on the USCS soil type yielded results on the order of the baseline naïve mean performance or slightly worse (Table 5.6), and the ranges captured fifty percent or less of the data for each soil type. A breakdown of these results by USCS soil type reveals that the error rate increases for soils with poorer gradation (GW and SW vs. GP and SP), decreasing particle size (gravels vs. sands), and greater amounts of plastic fines (silts and clays). Each of these contributing to greater unpredictability is expected,

and they agree with the finding that the coarse particle sizes are most highly correlated with strength. One exception to these trends is the result for the CL soil, which showed a relatively low error rate of sixty-seven to seventy-seven percent NRMSE. The reason for this is not clear, as we would expect a clay soil to exhibit a high level of variability and be more unpredictable. The result is based on a total of 192 cases in the database, the second highest of any soil type, so a small sample size is not a problem here.

The Mechanistic-Empirical Design Guide method for plastic soils gave a seventy-nine percent NRMSE on cases in the database, but for non-plastic soils the error rate was about fifty-two percent (Table 5.7). For both groups of soils, using the equivalent input features with a neural network model did not provide any appreciable improvement in performance.

Narrowing the scope of the analysis to a single soil type from one geographic location did not improve performance (Table 5.3). It was expected that better results would be possible for a more specific model. However, this may indicate that the general models designed to handle a wide variety of soils are performing as well as can be expected overall, based on the limitations of the dataset available for learning.

Because complex models appear to give no improvement for non-plastic soils, a basic approach is recommended. For example, the Mechanistic-Empirical Design Guide's method of using the 60^{th} percentile of the grain size to predict CBR is simple and provides results that are at least no worse than the baseline. A more complex model is appropriate for plastic soils. *k*-nearest neighbor with plasticity, moisture content, grain size, density and specific gravity as inputs attained the lowest error magnitude in this study and is recommended. If there are limitations of "portability" in applications where

the whole database cannot be available for the *k*-NN method predictions, then a neural network approach could be used with some minor loss of performance.

The lowest generalization error achieved with the approaches tested in the current research was approximately fifty percent. With the NRMSE representing one standard deviation of the residuals on a normalized basis, model predictions must be cut in half to provide comparable levels of confidence to those ordinarily used for fixed facilities. Possible causes of this large error include a dataset that does not cover the problem domain well enough, models that lack relevant input features to use for making good predictions, and noisy data. Because the database was deliberately assembled to avoid problems with the two former issues, it is more likely that the noise in both the target CBR parameter and several input features are causing the problems with prediction performance.

With such noisy measurements to work with, an obvious question is why not just select another index for soil strength that can be measured more precisely. Unfortunately, "the engineering properties of soil and rock exhibit varied and uncertain behavior due to the complex and imprecise physical properties associated with the formation of these materials" [11]. So, like many other geotechnical properties, soil strength is a tremendously variable property. The coefficient of variation reported for CBR in the literature of approximately twenty-five percent is sobering with respect to the Bayes error rate, as explained in section 5.2.1. In fact, *all* the standard methods that geotechnical engineers have available to measure soil strength appear to have comparable levels of inherent noise—generally on the order of ten to forty percent or more for natural soils [93]. This includes other field tests like shear vane, plate bearing subgrade

modulus, & deflectometer and even the more tightly controlled laboratory methods like triaxial, shear box, & unconfined compressive strength. Therefore, it is not likely that the high degree of variability is due to a lack of precision in the test apparatuses themselves, operator error, or interlaboratory differences in procedure, but caused by the intrinsic properties of the soil itself.

7 Recommendations for Future Work

Primarily due to the inherent variability of soil itself, it is very likely that soil strength prediction will remain a challenging problem for a long time to come. The difficulties encountered in trying to formulate improved techniques during the current research effort have revealed several areas where further efforts may prove the most fruitful. These include some trials with other prediction methods, strategies for future data collection efforts, and some thoughts regarding alternate approaches to the problem.

Additional Methods

Because of the relative compatibility of fuzzy logic's "Intelligence Density Profile" strengths in comparison with the constraints of the Opportune Landing Site problem, this technique should be paired with some of the prediction methods and tested. Hybridization of *k*-nearest neighbor or artificial neural networks with the representational flexibility of fuzzy set theory may lead to more accurate soil strength predictions. Other geotechnical applications have reported better performance of "neuro-fuzzy" approaches in comparison to conventional ANN models [134,135].

Another topic that deserves further exploration is the application of techniques that can deal with missing and categorical data more effectively. These may allow extra information to be gleaned from the current database and produce further insights into the problem.

Future Data Collection

Additional efforts to compile data for this approach to soil strength determination should concentrate on developing a database of laboratory measurements, which

presumably would be more closely-controlled than field testing. However, like the collection effort undertaken for this study, special care to gather data that spans a wide variety of soil types tested under conditions that are representative of those typically encountered in the field will be essential. With a laboratory dataset that contains less noise, hopefully some models could achieve performance levels closer to the fundamental limits posed by the intrinsic variability of the soil.

Alternative Approaches

Due to the difficulties encountered in predicting soil strength with real in situ measurements of key soil properties, the use of remote sensing signals to infer some of these physical properties and then use them as prediction model inputs does not seem very promising. This extra series of approximations will surely introduce even more error into a process already struggling to achieve satisfactory levels of precision. Instead of performing this conversion to intermediary soil properties, perhaps a testing program where the spectral signals and strength are both measured simultaneously across an area and the remote sensing data is used as a direct input for a prediction model might be a more fundamentally sound approach. Or, it is conceivable that the problem could be turned into a classification task if basic indicators that distinguish the spectral signatures of active landing zones from unsuitable areas could be reliably established.

In order to avoid the limitations posed by indirect measurement, the direct determination of soil properties is clearly more desirable from an error standpoint. However, implementing this in a manner that does not pose a risk to personnel is crucial. Perhaps some disposable miniaturized sensors could be dropped across unsurfaced areas of interest to measure critical geotechnical properties, including some direct assessment

of strength. There were some early attempts at developing air-dropped penetrometers to directly measure soil strength [136,137], but it may be time to revisit this approach given modern advancements in electronic sensor and networking technologies.

A. List of Abbreviations and Acronyms

#4 Avg	Average percent passing the number 4 sieve (4.75 mm)
#4 M	Maximum percent passing the number 4 sieve (4.75 mm)
#4 m	Minimum percent passing the number 4 sieve (4.75 mm)
#40 Avg	Average percent passing the number 40 sieve (425 µm)
#40 M	Maximum percent passing the number 40 sieve (425 µm)
#40 m	Minimum percent passing the number 40 sieve (425 µm)
#200 Avg	Average percent passing the number 200 sieve (75 µm)
#200 M	Maximum percent passing the number 200 sieve (75 µm)
#200 m	Minimum percent passing the number 200 sieve (75 µm)
0.005 Avg	Average percent finer than the 0.005 mm grain size
0.005 M	Maximum percent finer than the 0.005 mm grain size
0.005 m	Minimum percent finer than the 0.005 mm grain size
0.001 Avg	Average percent finer than the 0.001 mm grain size
0.001 M	Maximum percent finer than the 0.001 mm grain size
0.001 m	Minimum percent finer than the 0.001 mm grain size
3/4 Avg	Average percent passing the 3/4 inch sieve (19 mm)
3/4 M	Maximum percent passing the 3/4 inch sieve (19 mm)
3/4 m	Minimum percent passing the 3/4 inch sieve (19 mm)
AASHTO	American Association of State Highway and Transportation Officials
AFCEC	Air Force Civil Engineering Center
AFCESA	Air Force Civil Engineering Support Agency
AFESC	Air Force Engineering and Services Center
ANN	Artificial neural network
ASAE	American Society of Agricultural Engineers
ASCE	American Society of Civil Engineers
ASTM	American Society for Testing and Materials
CBR	California bearing ratio
CESC	Civil Engineering Division (AFCESA)
CI	Cone Index
CONUS	Continental United States
CRREL	Cold Regions Research and Engineering Laboratory
CV	coefficient of variation
DCP	Dynamic cone penetrometer
DD	Dry density
DELVE	Data for Evaluating Learning in Valid Experiments
DoD	Department of Defense
ERDC	Engineer Research and Development Center
FAO	Food and Agriculture Organization (of the United Nations)
FAMECE	Family of Military Engineer Construction Equipment
FASST	Fast All-season Soil STrength model
FFBP	Feed-forward back-propagation
ft	foot
ICAO	International Civil Aviation Organization
IDF	Intelligence Density Framework

IDP	Intelligence Density Profile
in	inch
ISO	International Standards Organization
ISRIC	International Soil Reference Information Center
JOA	Joint Operational Area
JRAC	Joint Rapid Airfield Construction
KBS	Knowledge-based system
***k*-NN**	*k*-nearest neighbor
lb	pound
LL	Liquid limit
MAF	Mobility Air Force
MC	Moisture content (gravimetric basis)
MDD	Maximum dry density
M-E	Mechanistic-Empirical
NCHRP	National Cooperative Highway Research Program
NN	Neural network
NRC	National Research Council
NRCS	Natural Resources Conservation Service
OCONUS	Outside Continental United States
OMC	Optimum moisture content
OLS	Opportune landing site
PAVER	Pavement Management System Software
PCA	Portland Cement Association
PCASE	Pavement-Transportation Computer Assisted Structural Engineering
PI	Plasticity index
PL	Plastic limit
psi	pounds per square inch
SINFERS	Soil Inference System
SMSP II	Soil Moisture Strength Prediction Model Version II
SpGr	Specific gravity
SPRO	Semi-Prepared Runway Operations
SSTOL	Super Short Takeoff and Landing
TRADOC	Training and Doctrine Command
USACE	United States Army Corps of Engineers
USAF	United States Air Force
USCS	Unified Soil Classification System
USDA	United States Department of Agriculture
WES	Waterways Experiment Station

B. Glossary of Selected Geotechnical Terms

These definitions are based on material drawn from several sources [20,138]. Also refer to database field descriptions in Appendix C for more definitions.

Base or Subbase Courses: Natural or processed materials placed on the subgrade beneath the pavement.

California Bearing Ratio (CBR): An empirical measure of soil strength used in the conventional design and evaluation of flexible pavement and unsurfaced airfields. To determine a CBR, a dynamic load is applied to a piston whose end is 3 square inches in area, forcing it to penetrate the soil at a rate of 0.05 inch/minute. The load required in pounds per square inch (psi) to force penetration gives the modulus of shear that is converted to a CBR using established load factors. Penetration into a crushed well-graded limestone serves as the benchmark material with a CBR of 100.

Compaction: The densification of a soil by means of mechanical manipulation.

Dynamic Cone Penetrometer (DCP): A probe-type instrument consisting of a cone-tipped rod that is driven into the soil by a sliding hammer. The DCP provides an indication of soil strength in terms of a DCP index. This index can then be used to estimate a CBR value.

Flexible Pavement: A pavement with a bituminous surface course and one or more supporting base or subbase courses placed over a prepared subgrade.

Liquid Limit: (1) The water content corresponding to the arbitrary limit between the liquid and plastic states of consistency of a soil. (2) The water content at which a pat of soil, cut by a groove of standard dimensions, will flow together for a distance of ½ inch under the impact of 25 blows in a standard liquid limit apparatus.

Maximum Dry Density: The dry unit weight defined by the peak of a compaction curve, as determined with a Proctor test.

Optimum Moisture Content: The water content at which a soil can be compacted to a maximum dry unit weight by a given compactive effort.

Plasticity Index: Numerical difference between the liquid limit and the plastic limit.

Plastic Limit: (1) The water content corresponding to an arbitrary limit between the plastic and semisolid states of consistency of a soil. (2) Water content at which a soil will just begin to crumble when rolled into a thread approximately ⅛ inch in diameter.

Proctor Test: A laboratory compacting procedure whereby a soil at a known water content is placed in a specified manner into a mold of a given dimensions, subjected to a compactive effort of controlled magnitude, and the resulting unit weight determined. The

procedure is repeated for various water contents sufficient to establish a relationship (known as a Proctor curve) between water content and unit weight. Also known as a **compaction test** or **moisture-density test**.

Rigid Pavement: A pavement consisting of a nonreinforced Portland cement concrete (PCC) surface course resting directly on a prepared subgrade, granular base course, or stabilized layer.

Subgrade: The natural in-place soil upon which a pavement, base, or subbase course is constructed.

Unified Soil Classification System (USCS): System developed by the U.S. Army Corps of Engineers (USACE) to group or classify soils based upon particle size, gradation, and plasticity characteristics, and rates their suitability as airfield construction materials.

C. Opportune Landing Site California Bearing Ratio Database Fields

N = numerical feature *C = categorical feature*
O = ordinal feature *B = binary feature*

OLS Data Point # *{N}*
Specific ID number given to each line of data as a unique identifier in the database.

JRAC Soil # *{N}*
Specific ID number given to each unique soil that was identified in the Joint Rapid Airfield Construction database.

Test or Sample Date *{N}*
Date on which measurements or tests were performed.

Report # *{C}*
Report Date *{N}*
Report Title *{C}*
Citation info for source of soil test data.

Country Code (ISO-3166) *{C}*
Standard two letter ID code for country in which test site is located [91].

Location *{C}*
Geographic location of test site (name of military base, town/state, airfield name, etc.).

Test Station *{C}*
Location or ID for test site within the geographic location given above (test pit #, location #, station on runway/taxiway, etc.).

Layer *{O}*
Layer in the pavement structure that the data has come from—used to distinguish engineered materials from more naturally occurring ones. *Base* (high quality material placed directly beneath the pavement), *Subbase* (lower quality select material placed below the base course) and *Subgrade* (natural soil found in place, may be compacted but otherwise unmodified).

Landform *{C}*
The category of landform based on slope, relief, and relation to surrounding lands for the general area surrounding the test site. Hierarchical categories based on [47], include:

- *L* *Level Land*
  - *LP* *Plains*
  - *LL* *Plateaux*
  - *LD* *Depressions*
  - *LF* *Low-gradient footslopes*
  - *LV* *Valley floors*

- *S* *Sloping Land*
  - *SM* *Medium-gradient mountains*
  - *SH* *Medium-gradient hills*
  - *SE* *Medium-gradient escarpment zone*
  - *SR* *Ridges*
  - *SU* *Mountainous highland*
  - *SP* *Dissected plains*

- *T* *Steep Land*
  - *TM* *High-gradient mountains*
  - *TH* *High-gradient hills*
  - *TE* *High-gradient escarpment zone*
  - *TV* *High-gradient valleys*

- *C* *Lands with Composite Landforms*
  - *CV* *Valleys*
  - *CL* *Narrow plateaux*
  - *CD* *Major depressions*

Lithology of Parent Material *{C}*
Category of rock type that forms the basis for the soil, primarily based on geology and mineralogy. Hierarchical categories based on [47], include:

I Igneous rock
- *IA Acid Igneous*
  - *IA1 Granite*
  - *IA2 Grano-Diorite*
  - *IA3 Quartz-Diorite*
  - *IA4 Rhyolite*
- *II Intermediate Igneous*
  - *II1 Andesite, Trachyte, Phonolite*
  - *II2 Diorite-Syenite*
- *IB Basic Igneous*
  - *IB1 Gabbro*
  - *IB2 Basalt*
  - *IB3 Dolerite*
- *IU Ultrabasic Igneous*
  - *IU1 Peridotite*
  - *IU2 Pyroxenite*
  - *IU3 Ilmenite, Magnetite, Ironstone, Serpentine*

M Metamorphic rock
- *MA Acid Metamorphic*
  - *MA1 Quartzite*
  - *MA2 Gneiss, Migmatite*
  - *MA3 Slate, Phyllite (peltic rocks)*
  - *MA4 Schist*
- *MB Basic Metamorphic*
  - *MB1 Slate, Phyllite (peltic rocks)*
  - *MB2 Schist*
  - *MB3 Gneiss rich in ferro-magnesian minerals*
  - *MB4 Metamorphic Limestone (Marble)*

S Sedimentary rock
- *SC Classic Sediments*
  - *SC1 Conglomerate, Breccia*
  - *SC2 Sandstone, Greywacke, Arkose*
  - *SC3 Siltstone, Mudstone, Claystone*
  - *SC4 Shale*
  - *SC5 Ironstone*
- *SO Organic*
  - *SO1 Limestone, other carbonate rocks*
  - *SO2 Marl and other mixtures*
  - *SO3 Coals, Bitumen, & related rocks*
- *SE Evaporites*
  - *SE1 Anhydrite, Gypsum*
  - *SE2 Halite*

Deposition Type *{C}*
Method of natural deposition for soil material at the test site. Categories for unconsolidated sediments based on [47], include:

UF	*Fluvial*
UL	*Lacustrine*
UM	*Marine*
UC	*Colluvial*
UE	*Eolian (Aeolian)*
UG	*Glacial*
UP	*Pyroclastic*
UO	*Organic*

Depth to Water Table *{N}*
Depth in feet to natural ground water from grade level at test site.

Soil Type, USCS *{C}*
Soil classification according to the Unified Soil Classification System. Twenty-six possible entries include:

GW	*Well-graded gravel*
GP	*Poorly-graded gravel*
GM	*Silty gravel*
GC	*Clayey gravel*
SW	*Well-graded sand*
SP	*Poorly-graded sand*
SM	*Silty sand*
SC	*Clayey sand*
ML	*Low-compressibility silt*
CL	*Lean clay*
OL	*Organic silt or clay*
MH	*High-compressibility silt*
CH	*Fat clay*
OH	*Organic silt or clay*
Pt	*Peat*
CL-ML	*Silty clay*
GW-GM	*Well-graded gravel with silt*
GW-GC	*Well-graded gravel with clay*
GP-GM	*Poorly-graded gravel with silt*
GP-GC	*Poorly-graded gravel with clay*
GC-GM	*Silty, clayey gravel*
SW-SM	*Well-graded sand with silt*
SW-SC	*Well-graded sand with clay*
SP-SM	*Poorly-graded sand with silt*
SP-SC	*Poorly-graded sand with clay*
SC-SM	*Silty, clayey sand*

Alternate Soil Type *{C}*
Alternate Soil System *{C}*
Soil classification with non-USCS system.

Soil Description *{C}*
Remarks on descriptive soil characteristics included with test data (textural description, color, etc.)

Clay Mineralogy *{C}*
The dominant type of mineral in the clay fraction of the soil. Can have a large influence on mechanical behavior for certain minerals. Categories based on [47], include:

AL	*Allophane*
CH	*Chloritic*
IL	*Illitic*
IN	*Interstratified or Mixed*
KA	*Kaolinitic*
MO	*Montmorillonitic*
SE	*Sesquioxidic*
VE	*Vermiculitic*

Specific Gravity *{N}*
Relative density of soil particles compared to water.

Sample Depth Below Grade *{N}*
Depth in inches from grade level at site where testing was performed.

Plastic or Non-Plastic *{B}*
Indicates whether a soil exhibits plastic behavior at some moisture content (e.g. clay) or does not (e.g. sand).

LL *{N}*
Liquid Limit of the soil in percent. The gravimetric moisture content at an arbitrary limit between the liquid and plastic states of consistency where the soil begins to exhibit a liquid behavior and will flow under its own weight.

PL *{N}*
Plastic Limit of the soil in percent. The gravimetric moisture content at an arbitrary limit between the plastic and semi-solid states of consistency where the soil begins to exhibit a plastic behavior and will deform under pressure without crumbling.

PI *{N}*
Plasticity Index of the soil in percent. The numerical difference between the liquid limit and plastic limit of the soil. A larger plasticity index indicates a soil that is more likely to exhibit plastic behavior.

Compactive Effort *{N}*
The amount of energy in foot-pounds per cubic foot put into compacting a unit volume of soil in preparing a laboratory sample. Different test standards result in different compactive efforts. Influences the shape and location of the compaction curve relating soil moisture to density.

Molding Moisture Content *{N}*
The gravimetric moisture content of the soil in percent used in preparing a laboratory sample.

Dry Density (laboratory) *{N}*
The density of the soil in pounds per cubic foot used in preparing a laboratory sample. The dry density includes only the oven-dry mass of soil particles present in a unit volume, not any of the free water that may exist contributing to the sample's moisture content.

Optimum Moisture Content and Max. Density *{B}*
An indication of whether the previous three measurements relate the peak on the moisture-density curve for that compaction energy (Y) or simply a single data point from a Proctor test on the moisture-density curve (N).

Unsoaked CBR (laboratory) *{N}*
Soaked CBR (laboratory) *{N}*
Laboratory measurement of the California bearing ratio in percent. The soil sample is prepared at a given compaction energy, molding moisture content, and dry density. It is then tested (unsoaked) or allowed to soak in water for four days to reach a nearly-saturated moisture condition.

Moisture Content as Tested (weight %) *{N}*
Moisture Content as Tested (volumetric %) *{N}*
The moisture content of the soil tested in percent. Gravimetric moisture content is the weight of free water in the soil that can be driven off by oven drying divided by the dry soil weight. Volumetric moisture content is the volume of water relative to the total volume of soil.

Trafficability Cone Index (CI) *{N}*
Index test of soil strength used for ground vehicle mobility. Performed by pushing a standard rod with 30º cone-shaped tip through the soil surface and recording the reaction force in psi. Test is performed on soil that is undisturbed.

Remolding Index *{N}*
A ratio of the trafficability cone index for undisturbed soils to those that have been remolded. This gives some indication of the change in vehicle mobility after many passes have occurred.

DCP Index (dynamic cone penetrometer) *{N}*
Dynamic cone penetrometer index test for soil strength, measured in millimeters per blow. Performed by using a sliding weight, repeatedly dropped from a constant height, to dynamically drive a 60º conically tipped rod through the soil. The distance of penetration is measured versus the number of blows and can be correlated with CBR.

Field CBR *{N}*
In situ field measurement of the California bearing ratio in percent.

Field Dry Density *{N}*
Field Wet Density *{N}*
The density of the soil measured in situ in the field. The dry density includes only the oven-dry mass of soil particles present in a unit volume—not any of the free water that may exist contributing to the sample's moisture content. The wet density includes both the oven-dry mass of soil particles present in a unit volume and any of the free water that may exist contributing to the sample's moisture content.

¾ inch Sieve, Maximum Percent Passing *{N}*
¾ inch Sieve, Minimum Percent Passing *{N}*
⅜ inch Sieve, Maximum Percent Passing *{N}*
⅜ inch Sieve, Minimum Percent Passing *{N}*
#4 Sieve, Maximum Percent Passing *{N}*
#4 Sieve, Minimum Percent Passing *{N}*
#10 Sieve, Maximum Percent Passing *{N}*
#10 Sieve, Minimum Percent Passing *{N}*
#40 Sieve, Maximum Percent Passing *{N}*
#40 Sieve, Minimum Percent Passing *{N}*
#100 Sieve, Maximum Percent Passing *{N}*
#100 Sieve, Minimum Percent Passing *{N}*
#200 Sieve, Maximum Percent Passing *{N}*
#200 Sieve, Minimum Percent Passing *{N}*
0.005 mm, Maximum Percent Passing *{N}*
0.005 mm, Minimum Percent Passing *{N}*
0.001 mm, Maximum Percent Passing *{N}*
0.001 mm, Minimum Percent Passing *{N}*
The gravimetric percentage of particles in a soil that are smaller than a certain size. Determined by shaking coarse soil particles through a stack of standard size sieves. For particles finer than the #200 sieve this is determined using a hydrometer by taking readings of a mixture of fine soil particles and water—with decreasingly smaller particles settling out of suspension over time. Both minimum and maximum are recorded due to soil data being grouped in many cases and the gradation plot resulting in a band of sizes rather than a distinct curve. If minimum equals maximum then data was recorded from a single curve (or converging band).

Roundness, Gravel *{N}*
Roundness, Sand *{N}*
Standard measure of the relative angularity of a soil particle's edges & corners, determined visually [44].

Sphericity, Gravel *{N}*
Sphericity, Sand *{N}*
Standard measure of the aspect ratio of a soil particle's dimensions, determined visually [44].

Remarks *{C}*
Catch-all for any remarks associated with test data.

Bibliography

[1] US Air Force, Air Mobility Command. *Air Mobility Master Plan 2004*. Scott AFB, Illinois: AMC, 2004.

[2] Sawyer, T. “High-Tech Tools and Hard, Hard Work at FOB Rhino.” *Engineering News-Record* 248.7 (25 February 2002): 20–22.

[3] US Army Training and Doctrine Command. *Force Operating Capabilities*. Pamphlet 525-66. Fort Monroe, Virginia: TRADOC, 1 July 2005.

[4] Stephenson, D. “Mobility Matters.” *Boeing Frontiers Magazine* February 2003.

[5] Vincent, R.K. and D.L. Jennings. “A Four-State Field Evaluation of the Boeing Landing Suitability Index (BLSI) for Automatically Mapping Candidate Aircraft Operating Sites in Natural Terrain from LANDSAT TM Data.” *Journal of Terramechanics* 41 (2004): 151–162.

[6] Berney, E.S. “Joint Rapid Airfield Construction: Site Selection and Assessment.” Presented at the *Joint Aerial Port of Debarkation Workshop, Destin, Florida, 29–31 March 2005.*

[7] Bouma, J. “Using Soil Survey Data for Quantitative Land Evaluation.” *Advances in Soil Science* 9 (1989): 177–213.

[8] Grunwald, S., D.J. Rooney, K. McSweeney, and B. Lowery. “Development of Pedotransfer Functions for a Profile Cone Penetrometer.” *Geoderma* 100 (2001): 25–47.

[9] McBratney, A.B., B. Minasny, S.R. Cattle, and R.W. Vervoort. “From Pedotransfer Functions to Soil Inference Systems.” *Geoderma* 109 (2002): 41–73.

[10] Van den Akker, J.J.H. “Development, Verification and use of the Subsoil Compaction Model SOCOMO.” *Proceedings of the 1st Workshop of the Concerted Action "Experiences with the Impact of Subsoil Compaction on Soil, Crop Growth and Environment and ways to Prevent Subsoil Compaction," Wageningen, Netherlands 28–30 May 1998.* Netherlands: DLO Winand Staring Centre, 1999. Report 168, pp. 321–336.

[11] Shahin, M.A., M.B. Jaska, and H.R Maier. “Artificial Neural Network Applications in Geotechnical Engineering.” *Australian Geomechanics* 36.1 (March 2001): 49–62.

[12] Toll, D.G. “Artificial Intelligence Systems for Geotechnical Engineering.” *Proceedings of the 10th European Young Geotechnical Engineers Conference, Izmir, Turkey, 21–24 October 1996*. Istanbul: Turkish National Committee for ISSMFE, 1996. pp. 211–238.

[13] Sarle, W.S., ed. Neural Network FAQ, periodic posting to the Usenet newsgroup comp.ai.neural-nets, Available online at <ftp://ftp.sas.com/pub/neural/FAQ.html>.

[14] US Army, Air Force, & Navy. *Materials Testing*. US Army Field Manual 5-530/Air Force AFM 89-3/Navy NAVFAC MO-330. Washington, DC: DoD, 17 August 1987.

[15] American Society for Testing and Materials. D 1883 – 05, Standard Test Method CBR (California Bearing Ratio) of Laboratory Compacted Soils. In *Annual Book of ASTM Standards*, Volume 04.08. West Conshohocken, Pennsylvania: ASTM, August 2005.

[16] American Society for Testing and Materials. D 4429 – 04, Standard Test Method CBR (California Bearing Ratio) of Soils in Place. In *Annual Book of ASTM Standards*, Volume 04.08. West Conshohocken, Pennsylvania: ASTM, 2004.

[17] US Department of Defense. *Test Methods for Pavement Subgrade, Subbase, and Base-Course Materials*. MIL-STD-621A. Washington, DC: DoD, 22 December 1964.

[18] Rollings, M.P., and R.S Rollings. G*eotechnical Materials in Construction*. New York: McGraw-Hill, 1996.

[19] Fine, L. and J.A. Remington. “Airfields for Very Heavy Bombers.” *United States Army in World War II - The Corps of Engineers: Construction in the United States*. Washington, DC: Office of the Chief of Military History, US Army, 1972. pp. 614–649.

[20] Air Force Civil Engineer Support Agency. *Airfield Pavement Evaluation: Standards and Procedures*. Engineering Technical Letter 02-19. Tyndall Air Force Base, Florida: AFCESA, October 2002.

[21] Rollings, R.S. “Evolution of Airfield Design Philosophies.” *Proceedings of the 22nd PIARC World Roads Congress, Durban, South Africa, 19–25 October 2003*. Cedex, France: World Road Association, 2003.

[22] Huang, Y.H. *Pavement Analysis and Design*. Englewood Cliffs, New Jersey: Prentice-Hall, 1993.

[23] Rollings, R.S. “RE: CBR/layered elastic discussions and Implications for Research.” E-mail to the author. 22 Nov. 2005.

[24] Gervais, E. “What Does an Aircraft Expect From the Pavement?” *Proceedings of the 7th International Conference on the Bearing Capacity of Roads, Railways, and Airfields, Trondheim, Norway, 27–29 June 2005*. Trondheim: Norwegian University of Science and Technology, 2005. Paper no. 285, 11 p.

[25] American Society for Testing and Materials. D 2487 – 00, Standard Practice for Classification of Soils for Engineering Purposes (Unified Soil Classification System). In *Annual Book of ASTM Standards*, Volume 04.08. West Conshohocken, Pennsylvania: ASTM, May 2000.

[26] Casagrande, A. “Classification and Identification of Soils.” *Transactions, American Society of Civil Engineers* 113 (1948): 901–991.

[27] US Army and Air Force. *Soils and Geology Procedures for Foundation Design of Buildings and Other Structures (Except Hydraulic Structures)*. US Army Technical Manual 5-818-1/Air Force AFM 88-3, Chapter 7. Washington, DC: US Army and Air Force, October 1983. Available online: < http://www.usace.army.mil/inet/usace-docs/armytm/tm5-818-1/ch-3.pdf>.

[28] Yoder, E.J. and M.W. Witczak. *Principles of Pavement Design*. 2nd ed. New York: John Wiley and Sons, 1975.

[29] Portland Cement Association. *PCA Soil Primer*. EB007.05S. Skokie, Illinois: PCA, 1992.

[30] US Army Corps of Engineers. *Characteristics of Soil Groups Pertaining to Roads and Airfields*. Appendix B, “The Unified Soil Classification System.” Technical Memorandum 3-357. Vicksburg, Mississippi: Waterways Experiment Station, 1960.

[31] US Army. *Military Soils Engineering*. Chapter 5, “Soil Classification.” Field Manual 5-410, Change 1. Washington, DC: US Army, June 1997.

[32] National Cooperative Highway Research Program. *Guide for Mechanistic-Empirical Design of New and Rehabilitated Pavement Structures. Final Report for Project 1-37A, Appendix CC-1: Correlation of CBR Values with Soil Index Properties*. Washington, DC: NCHRP, Transportation Research Board, National Research Council, 2004.

[33] Rada, G.R., C.W. Schwartz, M.W. Witczak, and S. Jafroudi. “Analysis of Climate Effects on Performance of Unpaved Roads.” *Journal of Transportation Engineering* 115.4 (1989): 389–410.

[34] Pavement Consultancy Services, Inc. *Damage Analysis of Unpaved Roads*. Technical Report prepared for The Earth Technology Corporation. Long Beach, California: 1987.

[35] US Army Corps of Engineers. *Soil Compaction Database.* Unpublished Compendium. Vicksburg, Mississippi: Waterways Experiment Station, 1999.

[36] US Army Corps of Engineers. *Tests on Soils Samples from Overseas Air Bases*. Technical Memorandum, No. 3-274. Vicksburg, Mississippi: Waterways Experiment Station, 1949.

[37] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Eielson Air Force Base, Alaska*. Tyndall Air Force Base, Florida: AFESC, February 1982.

[38] Attoh-Okine, B.N.O. "Modeling the Strength Characteristics of Lateritic Soils and Gravels in Pavement Construction." *Proceedings of the 4th International Conference on the Bearing Capacity of Roads and Airfields, Minneapolis, Minnesota, 17–21 July 1994*. Minneapolis: University of Minnesota, 1994. Volume 2, pp. 955–966.

[39] Attoh-Okine, N.O. and E.S.K. Fekpe. "Strength Characteristics Modeling of Lateritic Soils Using Adaptive Neural Networks." *Construction and Building Materials* 10.8 (1996): 577–582.

[40] Attoh-Okine, N.O. "Application of Genetic-based Neural Network to Lateritic Soil Strength Modeling." *Construction and Building Materials* 18.8 (2004): 619–623.

[41] Doshi, S.N. and H.R. Guirguis. "Correlation of CBR with Density and Moisture Content." *Indian Geotechnical Journal*. 12.4 (1983): 281–287.

[42] Black, W.P.M. "A Method of Estimating the California Bearing Ratio of Cohesive Soils from Plasticity Data." *Geotechnique* 12.4 (1962): 271–282.

[43] Kleven, J.R., T.B. Edil, and C.H. Benson. "Evaluation of Excess Foundry System Sands for Use as Subbase Material." *Transportation Research Record* 1714 (2000): 40–48.

[44] Krumbein, W. C., and Sloss, L. L. *Stratigraphy and Sedimentation*. San Francisco: W. H. Freeman, 1951.

[45] FAO-UNESCO. *FAO/UNESCO Soil Map of the World 1:5,000,000*. Paris: UNESCO, 1974.

[46] International Soil Reference and Information Center. *Soil and Terrain Database (SOTER)*. Wageningen, Netherlands: ISRIC, 2004. <http://www.isric.org/isric/webdocs/Docs/soterfolder001.pdf>

[47] van Engelen, V.W.P., and T.T. Wen, eds. *Global and National Soils and Terrain Digital Databases (SOTER): Procedures Manual (Revised Edition)*. Wageningen, Netherlands: ISRIC, 1995.

[48] American Society of Agricultural Engineers. S313.3, *Soil Cone Penetrometer*. St. Joseph, Michigan: ASAE, February 2004.

[49] Willoughby, W.E. *Ride test results for the FAMECE dumper and comparison vehicles*. Technical report GL-81-1. Vicksburg, Mississippi: Waterways Experiment Station, January 1981.

[50] Frankenstein, S. and G.G. Koenig. *Fast All-season Soil STrength*. ERDC/CRREL Special Report 04-1. Hanover, New Hampshire: ERDC/CRREL, September 2004.

[51] Sullivan, P.M., C.D. Bullock, N.A. Renfroe, M.R. Albert, G.G. Koenig, L. Peck, and K. O'Neill. *Soil Moisture Strength Prediction Model Version II (SMSP II)*. Technical Report GL-97-15. Vicksburg, Mississippi: Waterways Experiment Station, August 1997.

[52] SoilVision Systems Ltd. *SoilVision*, Version 4.02.006. Saskatoon, Saskatchewan: SoilVision Systems, 2005.

[53] Soil Survey Staff 2006. *National Soil Survey Characterization Data.* 2006. Soil Survey Laboratory, National Soil Survey Center, USDA-NRCS, Lincoln, Nebraska. 7 April 2006. < http://ssldata.nrcs.usda.gov/default.htm>.

[54] Dyke P.T., J.D. Nichols, and B. McMahon. *Global Soils Database Progress Report & Soils Boundary Enhancement Feasibility Study*. Washington: National Imaging and Mapping Agency, October 2003.

[55] Van Steenburg, G. "PCASE-PAVER-GeoBase Integration." Presented at the *Tri-Service Pavement Conference, Port Hueneme, California, 6–9 Dec 2004*.

[56] Toll, D.G. *GeotechML at University of Durham*. 2005. 30 August 2005. <http://www.dur.ac.uk/~des0www4/geotechml/#Links>.

[57] US Army Corps of Engineers. *Field Moisture Content Investigation*. Technical Memorandum No. 3-401, Report No. 2, October 1945 - November 1952 Phase. Vicksburg, Mississippi: Waterways Experiment Station, April 1955.

[58] Tingle, Jeb S. *Testing and Analysis of C-17 Live-Flight Operations on Semi-Prepared Airfields*. Technical Report GL-98-11. Vicksburg, Mississippi: Waterways Experiment Station, June 1998.

[59] Air Force Civil Engineering Support Agency. *(CESC) Pavement Reports - HQ AFCESA*. 2005. Tyndall Air Force Base, Florida. March 2005. Access restricted to .mil domains. <https://wwwmil.afcesa.af.mil/Directorate/CES/Civil/Pavements/pav_app/Pav_main.asp>.

[60] Air Force Civil Engineering Center. *Airfield Pavement Evaluation Report, Holloman Air Force Base, New Mexico*. Tyndall Air Force Base, Florida: AFCEC, October 1977.

[61] Air Force Civil Engineering Center. *Airfield Pavement Evaluation Report, Sondrestrom Air Base, Greenland*. Tyndall Air Force Base, Florida: AFCEC, December 1974.

[62] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Andersen Air Force Base, Guam*. Tyndall Air Force Base, Florida: AFESC, March 1985.

[63] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Hickam Air Force Base, Hawaii*. Tyndall Air Force Base, Florida: AFESC, August 1980.

[64] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Kingsley Air National Guard Base, Oregon*. Tyndall Air Force Base, Florida: AFESC, December 1987.

[65] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Quonset State Airport, Rhode Island*. Tyndall Air Force Base, Florida: AFESC, September 1981.

[66] Air Force Engineering and Services Center. *Partial Airfield Pavement Evaluation, Westover AFB, Massachusetts*. Tyndall Air Force Base, Florida: AFESC, December 1980.

[67] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, RAF Mildenhall AB, England*. Tyndall Air Force Base, Florida: AFESC, December 1982.

[68] Air Force Civil Engineering Center. *Airfield Pavement Evaluation and Condition Survey Report, Luke Air Force Base, Arizona*. Tyndall Air Force Base, Florida: AFCEC, May 1974.

[69] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, George Air Force Base, California*. Tyndall Air Force Base, Florida: AFESC, May 1983.

[70] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Wake Island*. Tyndall Air Force Base, Florida: AFESC, January 1980.

[71] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, McChord Air Force Base, Washington*. Tyndall Air Force Base, Florida: AFESC, June 1981.

[72] Air Force Civil Engineering Center. *Airfield Pavement Evaluation, Myrtle Beach Air Force Base, South Carolina*. Tyndall Air Force Base, Florida: AFCEC, August 1975.

[73] US Army Corps of Engineers. *Airfield Pavement Evaluation Report, Spangdahlem AB, Germany*. Cincinnati, Ohio: USACE Ohio River Division Laboratory, June 1969.

[74] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Castle AFB, California*. Tyndall Air Force Base, Florida: AFESC, December 1981.

[75] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Kadena Air Base, Okinawa, Japan*. Tyndall Air Force Base, Florida: AFESC, January 1985.

[76] Air Force Engineering and Services Center. *Partial Airfield Pavement Evaluation, Loring Air Force Base, Maine*. Tyndall Air Force Base, Florida: AFESC, December 1981.

[77] Air Force Engineering and Services Center. *Special Study, Airfield Pavement Evaluation, Marana Air Park, Arizona*. Tyndall Air Force Base, Florida: AFESC, November 1980.

[78] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Wheeler Air Force Base, Hawaii*. Tyndall Air Force Base, Florida: AFESC, August 1981.

[79] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Zaragoza AB, Spain*. Tyndall Air Force Base, Florida: AFESC, February 1984.

[80] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Indian Springs Air Force Auxiliary Field, Nevada*. Tyndall Air Force Base, Florida: AFESC, January 1982.

[81] Air Force Civil Engineering Center. *Airfield Pavement Evaluation, McGuire Air Force Base, New Jersey*. Tyndall Air Force Base, Florida: AFCEC, September 1978.

[82] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Nellis Air Force Base, Nevada*. Tyndall Air Force Base, Florida: AFESC, February 1982.

[83] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Clark AB, Philippines*. Tyndall Air Force Base, Florida: AFESC, April 1984.

[84] Air Force Engineering and Services Center. *Partial Airfield Pavement Evaluation, Hickam Air Force Base, Hawaii*. Tyndall Air Force Base, Florida: AFESC, October 1987.

[85] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Sidi Slimane AB, Morocco*. Tyndall Air Force Base, Florida: AFESC, March 1988.

[86] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Maxwell Air Force Base, Alabama*. Tyndall Air Force Base, Florida: AFESC, September 1980.

[87] Air Force Civil Engineering Center. *Airfield Pavement Evaluation Report, Eglin Air Force Base, Florida*. Tyndall Air Force Base, Florida: AFCEC, March 1975.

[88] Air Force Civil Engineering Center. *Airfield Pavement Evaluation and Condition Survey Report, George AFB, California*. Tyndall Air Force Base, Florida: AFCEC, February 1976.

[89] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Hancock Field, New York*. Tyndall Air Force Base, Florida: AFESC, February 1980.

[90] Air Force Engineering and Services Center. *Airfield Pavement Evaluation, Kadena Air Base, Okinawa, Japan*. Tyndall Air Force Base, Florida: AFESC, June 1979.

[91] International Standards Organization. ISO 3166, *English country names and code elements*. Geneva, Switzerland: ISO, 2005.

[92] Robinson, J.H. and C.P. Rabalais. *Performance of the Combat Engineer Vehicle with Mineplow Operating Worldwide and in Theaters of Operation*. Technical Report GL-93-23. Vicksburg, Mississippi: USACE Waterways Experiment Station, August 1993.

[93] Freeman, R.B. and W.P. Grogan. *Statistical Analysis and Variability of Pavement Materials*. Technical Report GL-97-12. Vicksburg, Mississippi: Waterways Experiment Station, August 1997.

[94] Wysocki, D.A., P.J. Schoeneberger, and H.E. LaGarry. “Geomorphology of Soil Landscapes.” Section E, Chapter 1. In *Handbook of Soil Science*. Ed. M.E. Sumner. Boca Raton, Florida: CRC Press, 2000. E1–E40.

[95] Brown, M.L. and J.F. Kros. “Data Mining and the Impact of Missing Data.” *Industrial Management & Data Systems*. 103.8 (2003): 611–621.

[96] Viharos, Z.J., L. Monostori, and T. Vincze. "Training and Application of Artificial Neural Networks with Incomplete Data." In *Lecture Notes of Artificial Intelligence, LNAI 2358, The Fifteenth International Conference on Industrial & Engineering Application of Artificial Intelligence & Expert Systems, Cairns, Australia, 17–20 June 2002*. Heidelberg, Germany: Springer-Verlag, 2002. pp. 649–659.

[97] SAS Institute Inc. *JMP*, Release 5.1.2. Cary, North Carolina: SAS, 2004.

[98] Masters, T. *Neural, Novel, & Hybrid Algorithms for Time Series Prediction*. New York: John Wiley and Sons, 1995.

[99] Dhar, V. and R. Stein. *Intelligent Decision Support Methods: The Science of Knowledge Work*. Upper Saddle River, New Jersey: Prentice-Hall, 1997.

[100] Michie, D., D.J. Spiegelhalter, and C.C. Taylor (eds). *Machine Learning, Neural and Statistical Classification*. Chichester, England: Ellis-Horwood, 1994.

[101] Haykin, S. *Neural Networks: A Comprehensive Foundation*. 2nd ed. Upper Saddle River, New Jersey: Prentice-Hall, 1999.

[102] Weiss, S.M. and C.A. Kulikowski. *Computer Systems that Learn*. San Mateo, California: Morgan Kaufmann, 1991.

[103] Newman, D.J., S. Hettich, C.L. Blake, and C.J. Merz. *UCI Repository of Machine Learning Databases*. 1998. Irvine, CA: University of California, Department of Information and Computer Science. 18 January 2006. <http://www.ics.uci.edu/~mlearn/MLRepository.html>.

[104] Rasmussen, C.E., R.M. Neal, G. Hinton, D. van Camp, M. Revow, Z. Ghahramani, R. Kustra, and R. Tibshirani. *The DELVE Manual*, Version 1.1. 1996. University of Toronto. 18 January 2006. < http://www.cs.toronto.edu/~delve>

[105] Juang, C.H., J.E. Clark, and P. Ghosh. "Representation, Processing, and Interpretation of Fuzzy Information in Civil Engineering." *Transportation Research Record* 1399 (1993): 20–26.

[106] Si, B.C. and W. Bodhinayake. "Determining Soil Hydraulic Properties from Tension Infiltrometer Measurements: Fuzzy Regression." *Soil Science Society of America Journal*. 69 (2005): 1922–1930.

[107] Mendel, J.M. *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions*. Upper Saddle River, New Jersey: Prentice-Hall, 2001.

[108] Jang, J.S.R., C.-T. Sun, and E. Mizutani. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence*. Upper Saddle River, New Jersey: Prentice-Hall, 1997.

[109] Khan, A.H. "On the Selection of a Suitable AI Technique for Solving a Given Problem." Tutorial. Presented at the *3rd IEEE National Multitopic Conference (INMIC'98), Lahore, Pakistan, 20-22 November 1998.* <http://www.altafkhan.com/pdf/ai-technique-selection.pdf>

[110] StatSoft, Inc. *Electronic Statistics Textbook*. Tulsa, Oklahoma: StatSoft, 2006. 31 January 2006. <http://www.statsoft.com/textbook/stathome.html>.

[111] Fayyad, U. G. Piatetsky-Shapiro, and P. Smyth. "From Data Mining to Knowledge Discovery in Databases." *AI Magazine* 17.3 (1996): 37–54.

[112] Montgomery, D.C., and G.C. Runger. *Applied Statistics and Probability for Engineers*. New York: John Wiley and Sons, 1994.

[113] Shahin, M.A., M.B. Jaksa, and H.R. Maier. "Artificial Neural Network Applications in Geotechnical Engineering." *Australian Geomechanics* 36.1 (2001): 49–62.

[114] Demuth, H., M. Beale, and M. Hagan. *Neural Network Toolbox For Use with MATLAB®: Users Guide, Version 4*. Natick, Massachusetts: The MathWorks, 2005.

[115] Cybenko, G.V. "Approximation by Superpositions of a Sigmoidal function." *Mathematics of Control, Signals and Systems*. 2 (1989): 303–314.

[116] Cybenko, G.V. *Continuous Valued Neural Networks with Two Hidden Layers Are Sufficient*. Technical Report. Medford, Massachusetts: Department of Computer Science, Tufts University, 1988.

[117] Heath, G. *How Many Hidden Neurons???* Posted to the Usenet newsgroup comp.ai.neural-nets. 1 December 2003. <http://groups.google.com/group/comp.ai.neural-nets/browse_frm/thread/d699e17a4cc926e2>.

[118] Sarle, W.S. *How to measure importance of inputs?* Revised 23 June 2000. 12 April 2006. <ftp://ftp.sas.com/pub/neural/importance.html >.

[119] Nauck, D., F. Klawonn, and R. Kruse. *Foundations of Neuro-Fuzzy Systems*. New York: John Wiley & Sons, 1997.

[120] Al-Amoudi, O.S.B., I.M. Asi, H.I.A. Wahhab, and Z.A. Khan. “Clegg Hammer—California-Bearing Ratio Correlations.” *Journal of Materials in Civil Engineering* 14.6 (2002): 512–523.

[121] The MathWorks Inc. *MATLAB*®, Version 7.1.0.246 (R14) Service Pack 3. Natick, Massachusetts: The MathWorks, 2 August 2005.

[122] Quinlan, J.R. “Combining Instance-Based and Model-Based Learning.” *Proceedings of the Tenth International Conference on Machine Learning, Amherst, Massachusetts, 27–29 June 1993*. San Mateo, California: Morgan Kaufmann, 1993. pp. 236–243.

[123] Heath, G. *training a neural network*. Posted to the Usenet newsgroup comp.ai.neural-nets. 19 January 2006. < http://groups.google.com/group/comp.soft¯sys.matlab/browse_frm/thread/59f9c92200da939>.

[124] Duda, R.O., P.E. Hart, D.G. Stork. *Pattern Classification*. 2nd ed. New York: John Wiley & Sons, 2001.

[125] Pittman, D.W. *Development of a Reliability-Based Design Procedure for Rigid and Flexible Airfield Pavements*. Miscellaneous Paper GL-96-8. Vicksburg, Mississippi: Waterways Experiment Station, April 1996.

[126] Hampton, D., E.J. Yoder, and I.W. Burr. “Variability of Engineering Properties of Brookston and Crosby Soils.” *Proceedings, Highway Research Board* 41 (1962): 621–649.

[127] National Research Council. *The AASHTO Road Test Report 6: Special Studies*. Special Report 61F, Publication 955. Washington, DC: NRC, 1962.

[128] Patterson, W.D.O. “State-of-the-Art in a Probabilistic Approach to the Analysis, Design and Performance of Pavement Systems.” *Proceedings of the Third International Conference on Applications of Statistics and Probability in Soil and Structural Engineering, Sydney, Australia, 29 January–2 February 1979*. Kensington, New South Wales: Unisearch, Ltd., 1979.

[129] Yeo, R.E.Y. and A.A. Auff. “Pavement Construction Variability—Local Roads.” *Proceedings of the Second International Conference on Road and Airfield Pavement Technology, Singapore, 27–29 September 1995*. Singapore: Center for Transportation Research, National University of Singapore, 1995. pp. 734–745.

[130] Mitchell, M.F., C.J. Semmelink, and A.L. McQueen. “Statistical Quality Assurance in Highway Engineering in South Africa.” *Transportation Research Record* 652 (1977): 58–65.

[131] Liu, T.K., and M.R. Thompson. "Variability of Some Selected Laboratory Soil Tests." *Proceedings of the National Conference on Statistical Quality Control Methodology in Highway and Airfield Construction, Charlottesville, Virginia, 3–5 May 1966*. Charlottesville: University of Virginia, November 1966. pp. 375–395.

[132] Noureldin, A.S., E. Sharaf; A. Arafah, and F. Al-Sugair. "Estimation of Standard Deviation of Predicted Performance of Flexible Pavements using AASHTO Model." *Transportation Research Record* 1449 (1994): 46–56.

[133] US Army Corps of Engineers. *Pavement Design for Airfields*. Unified Facilities Criteria 3-260-02. Washington, DC: USACE, 30 June 2001.

[134] Juang, C.H., S.H. Ni, and P.C. Lu. "Training Artificial Neural Networks with the Aid of Fuzzy Sets." *Computer-Aided Civil and Infrastructure Engineering* 14 (1999): 407–415.

[135] Kumar, J.K., M. Konno, and N. Yasuda. "Subsurface Soil-Geology Interpolation Using Fuzzy Neural Network." *Journal of Geotechnical and Geoenvironmental Engineering* 126.7 (2000): 632–639.

[136] US Army Corps of Engineers. *Studies of Aerial Cone Penetrometer. Report 1, Laboratory Study of Mechanical Principles*. No. 3-462. Vicksburg, Mississippi: Waterways Experiment Station, 1957.

[137] Marten, V.R. and V. Hansen. *An Inexpensive Aerial Penetrometer*. AFWL-TR-74-56. Albuquerque, New Mexico: Sandia Laboratory, 1976.

[138] American Society for Testing and Materials. D 653 – 97, Standard Terminology Relating to Soil, Rock, and Contained Fluids. In *Annual Book of ASTM Standards*, Volume 04.08. West Conshohocken, Pennsylvania: ASTM, August 1997.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
July 2006

2. REPORT TYPE

3. DATES COVERED *(From - To)*

4. Title And Subtitle

A Generalized Approach to Soil Strength Prediction with Machine Learning Methods

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Peter M. Semen

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Cold Regions Research and Engineering Laboratory
U.S. Army Engineer Research and Development Center
72 Lyme Rd.
Hanover, NH 03755

8. PERFORMING ORGANIZATION REPORT NUMBER

ERDC/CRREL TR-06-15

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Current methods for evaluating the suitability of potential landing sites for fixed-wing aircraft require a direct measurement of soil bearing capacity. In contingency military operations, the commitment of ground troops to carry out this mission prior to landing poses problems in hostile territory, including logistics, safety, and operational security. Developments in remote sensing technology provide an opportunity to make indirect measurements that may prove useful for inferring basic soil properties. However, methods to accurately predict strength from other fundamental geotechnical parameters are lacking, especially for a broad range of soil types under widely varying environmental conditions. To support the development of new procedures, a dataset of in situ soil pit test results was gathered from airfield pavement evaluations at 46 locations worldwide that encompass a broad variety of soil types. Many features associated with soil strength—including gradation, moisture content, density, specific gravity and plasticity—were collected along with California bearing ratio (CBR), a critical strength index used to determine the traffic loading that the ground can support. Machine learning methods—with advantages in nonlinear relationship mapping, nonparametric distribution treatment, superior generalization, and implicit modeling—were applied, hypothesizing these characteristics might make them better suited to geotechnical problems. Artificial neural network and *k*-nearest neighbor techniques were tested on plastic and non-plastic subsets of data and compared to conventional regression and existing CBR prediction methods. The machine learning models were able to halve the baseline error rate for plastic soils, but non-plastic soils showed no significant improvement. For both groups, normalized root mean square error (NRMSE) for generalization to new cases was approximately fifty percent for the best models. The high degree of variability for direct soil strength measurement methods limits the lowest possible NRMSE to approximately 25%, even before introducing any additional errors expected with remote sensing.

15. SUBJECT TERMS

Aircraft, fixed wing
Artificial neural network
Bearing capacity
Remote sensing
Soils

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(include area code)*
U	U	U	U	152	